Investigation of Mammalian Chromatin Folding at Different Genomic Length Scales Using High Resolution Imaging

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

im Fach Biologie

eingereicht an der
Lebenswissenschaftlichen Fakultät der
Humboldt-Universität zu Berlin

von

Dipl. Chem. Dorothee Charlotte Agathe Krämer

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät
Prof. Dr. Bernhard Grimm

Gutachter/innen 1. Prof. Dr. Ana Pombo

2. Prof. Dr. Petra Knaus

3. Prof. Dr. Markus Landthaler

Tag der mündlichen Prüfung: 12.11.2018

Erklärung

Hiermit erkläre ich, die Dissertation selbstständig und nur unter Verwendung der angegebenen Hilfen und Hilfsmittel angefertigt zu haben. Ich habe mich anderwärts nicht um einen Doktorgrad beworben und besitze keinen entsprechenden Doktorgrad. Ich erkläre, dass ich die Dissertation oder Teile davon nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht habe und dass sie dort weder angenommen noch abgelehnt wurde. Ich erkläre die Kenntnisnahme der dem Verfahren zugrundeliegenden Promotionsordnung der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin vom 5. März 2015. Weiterhin erkläre ich, dass keine Zusammenarbeit mit gewerblichen Promotionsberaterinnen/Promotionsberatern stattgefunden hat und dass die Grundsätze der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

Declaration

I hereby declare that I completed the doctoral thesis independently based on the stated resources and aids. I have not applied for a doctoral degree elsewhere and do not have a corresponding doctoral degree. I have not submitted the doctoral thesis, or parts of it, to another academic institution and the thesis has not been accepted or rejected. I declare that I have acknowledged the Doctoral Degree Regulations, which underlie the procedure of the Faculty of Life Sciences of Humboldt-Universität zu Berlin, as amended on 5th March 2015. Furthermore, I declare that no collaboration with commercial doctoral degree supervisors took place, and that the principles of Humboldt-Universität zu Berlin for ensuring good academic practice were abided by.

Berlin, ...

Dorothee Charlotte Agathe Krämer

Zusammenfassung

Chromatin ist ein Makromolekül, dessen Genregulation innerhalb des räumlich eingeschränkten Zellkerns organisiert werden muss. Die Balance, die dabei zwischen DNA-Struktur und DNA-Funktion entsteht, ist essentiell für die Zelle, da die hierdurch gewährleistete stabile Genregulation Entwicklung und Krankheit steuert. Die Chromatinorganisation ist somit eng mit Genaktivierung und Genrepression verbunden.

In den vergangenen Jahren wurden viele Fortschritte in der Erforschung der Chromatinarchitektur erreicht. Unter Verwendung von Mikrokopie, Chromsome Conformation Capture-Methoden, computergestütztem Modellieren und Genome Architecture Mapping (GAM) wurde gezeigt, dass die DNA hierarchisch organisiert ist. Die Faltung läuft in aufeinander folgenden Schritten ab, wobei jede Organisationsebene sowohl zur räumlichen Komprimierung, als auch zur Genregulation beiträgt.

Mit Hilfe von hochauflösender Mikroskopie habe ich verschiedene Ebenen der 3D Chromatinorganisation auf Einzelzell-Basis untersucht. Auf der kleinsten Organisationsebene habe ich die Struktur zweier, nebeneinander liegender topologischer Domänen (TADs) am Sox9-Lokus erforscht. Mit Hilfe von Fluoreszenz in situ Hybridisierung (FISH) in 3D Zellen, sowie Cryoschnitten in embryonalen Stammzellen von Mäusen konnte ich Interaktionen zwischen den benachbarten TADs feststellen. Indem ich FISH in Zellen mit genomischen Duplikationen durchführte, konnte ich ferner zwei unterschiedliche, durch die Duplikation entstandene, Konformationen nachweisen.

Neue Ergebnisse aus GAM Experimenten zeigten die Existenz von long-range, multimeren Kontakten zwischen regulatorischen Elementen. Ich untersuchte die Bildung dieser long-range Kontakte zwischen TADs in einzelnen Zellen mit Hilfe von FISH und konnte ein häufiges Auftreten dieser Kontakte zwischen regulatorischen Domänen nachweisen. Zudem konnte ich demonstrieren, dass sich Cluster zwischen mehreren, weit auseinander liegenden, regulatorischen Elementen bilden. Dies lässt unter Umständen auf das Entstehen von regulatorischen Zentren zwischen diesen Enhancer-reichen Regionen schließen. Weitere Untersuchungen zeigten die Veränderung der sogenannten Super-Enhancer Cluster in unterschiedlichen Zelltypen. Des Weiteren konnte ich nachweisen, dass die Super-Enhancer-

enthaltenden Zellregionen sehr dekondensiert sind und sich bevorzugt an Splicing-Speckle-Regionen anschließen.

Abstract

Chromatin needs to organize gene regulation whilst fitting into the confined space of the nucleus. Maintaining a balance between DNA structure and DNA function is crucial for the cell, as ensuring stable gene regulation is fundamentally important in development and disease. Chromatin organization is therefore intertwined with gene activation and silencing.

In recent years many advances in the field of chromatin architecture have been made. The use of Imaging based methods, Chromosome Conformation Capture techniques, computational modelling and Genome Architecture Mapping (GAM), have painted a picture where chromatin is organized hierarchically. Folding occurs in subsequent units, where each level of organization contributes to the spatial compaction of DNA and gene regulation.

Using single-cell, high-resolution imaging, I investigated different levels of 3D chromatin organization. On the smallest scale, I looked at the 3D organization of two neighbouring Topologically Associating Domains (TADs) at the Sox9 locus, which are separated by a TAD boundary. Performing *Fluorescence in situ Hybridization* (FISH) in 3D and cryosectioned mouse embryonic stem cells, I detected extensive contacts between the two neighbouring TADs. Applying FISH in a cell line bearing a genomic duplication within the Sox9 locus, I also detected the occurrence of two different conformations that result from the duplication.

Recent evidence from GAM showed the formation of long-range, multimer contacts between distal regulatory elements. I investigated the occurrence of long-range contacts between super-enhancer TADs in single cells by FISH, and showed that they establish frequent interactions at close spatial distances. I showed the formation of clusters containing distal super-enhancer TADs, indicating the possibility of higher-order regulatory hubs between these enhancer-rich regions. Further investigation showed that super-enhancer regions form different clusters in different cell types. Finally, I showed that super-enhancers are highly decondensed and preferentially located at splicing speckles.

Acknowledgements

First and foremost I want to thank my supervisor Ana Pombo for giving me the opportunity to work in her group. Thank you very much for your supervision, scientific discussions and support throughout the years. Thank you also, for letting me discover the beautiful and fascinating world of microscopy and image analysis.

Thank you Sheila for teaching me the preparation of cryosections, as well as how to perform cryoFISH and immunofluorescence stainings. Having worked with you during the first months of my PhD was invaluable. Thank you also for your all the advice throughout the years.

I want to thank all my past and present colleagues, the best possible colleagues I could have hoped for. Thank you Ines and Joao for being wonderful friends since the very start. Thank you Elena, for all the good times we had during our joint PhD journey. Thank you Tiago for support with Macro writing, introducing me to writing R and the piano fun. Thank you Gesa and Rieke, you girls bring glitter to life inside and outside of the lab. Thank you Sasha for all the help throughout the years. A big thanks goes to all my past and present colleagues from the Pombo Combo: Mariano, Giulia, Sasha, Rob, Markus, Anita, Iza, Tom, Carmelo, Marta, Eshan, Franka, Nadina and Kelly, thank you all for your friendship, help and creating a great atmosphere!

Many, many thanks also to Marta M., for everything! Thank you Till for your friendship and all the support in image analysis with writing scripts and the many explanations. I also thank all my wonderful friends who made the last years a great time. Thank you Astrid, having you in Berlin was a gift. Thank you Katharina and Steffi for all the memories we made.

Most importantly, I thank my parents and my family for their tireless support, love and understanding. For sharing with me moments of happiness and always being there in difficult ones. To you I dedicate this work.

"Nature uses only the longest threads to weave her patterns,
so that each small piece of her fabric reveals
the organization of the entire tapestry."

Richard Feynman

Table of Contents

Index of Figures

FIGURES CHAPTER 1

FIGURES CHAPTER 2

FIGURES CHAPTER 3

FIGURES CHAPTER 4

Index of Tables

Abbreviations

3C	Chromosome Conformation Capture
3D	three dimensional
Bp	base pairs
BAC	Bacterial Artificial Chromosomes
BET	Bromodomain and extra-terminal domain family of proteins
BSA	Bovine serum albumin
BRD4	bromodomain-containing protein 4
cryoFISH	FISH on cryosections
cryoIF	immunofluorescence on cryosections
CTCF	CCCTC-binding factor
CTD	carboxy-terminal domain
CAT	conserved associated topologies
cHi-C	capture Hi-C
DNA	deoxyribose nucleic acid DNase deoxyribose
DAPI	4',6-Diamidino-2-Phenylindole, Dihydrochloride
ESC	embryonic stem cell
HEPES	4-(2-hydroxyethyl)-1-piperazineethaneszlfonic acid
FISH	fluorescence *in situ* hybridization
GRO-seq	global run-on and sequencing
HCl	hydrochloric acid
GAM	genome architecture mapping
K	Lysine
Kb	kilobases
LAD	lamina-associated domain
Mb	megabases
mRNA	messenger RNA
PBS	phosphate buffered saline
PCA	principal components analysis
PFA	paraformaldehyde
RNA	ribose nucleic acid
RNAPII	RNA polymerase II

SE	super-enhancer
Seq	sequencing
SLICE	statistical inference of co-segregation
SSC	saline-sodium citrate
STED	stimulated emission depletion
SVs	structural variations
TAD	topologically associating domain
TF	transcription factor
tRNA	transfer RNA
WGA	whole genome amplification

1 Introduction

1.1 Genome architecture is at the core of cell identity

At every moment during an organism's lifetime, the DNA of each cell needs to orchestrate a multitude of transcriptional outputs that are crucial for its development and function. The way in which the DNA 3-dimensional (3D) organization enables a stable and efficient read-out of its genomic content lies at the heart of the field of genome architecture. Initially, it was thought that the DNA sequence code itself provides a blueprint of cellular transcription, and with sufficient knowledge the DNA could be "read" like a book. However, we now know that there is information beyond the DNA sequence, which is equally important in encoding the instructions for the development of an organism and its cellular homeostasis. Epigenetic control of gene activity is regulated by many mechanisms, including direct chemical modifications of the DNA sequence and modifications of histone proteins, which can initiate and/or maintain activation or repression of genes. Genome architecture is another essential factor organizing gene regulation. 3D Genome organization contributes to transcription regulation by spatially organizing transcription and silencing in many different ways. The spatial organization of the DNA is important because the nucleus is a confined environment and chromatin organisation needs to be highly functional to minimize the space occupied by the DNA whilst maintaining efficient, stable gene expression and allowing for gene regulation.

1.2 Gene regulation in mammalian cells

RNA polymerase II

RNA polymerases are enzymes that are responsible for transcribing DNA into RNA copies. In eukaryotes, RNA polymerase I (RNAPI) transcribes ribosomal RNA (rRNA) genes (45S pre-ribosomal RNA). RNA polymerase II (RNAPII) transcribes all protein coding genes into messenger RNAs (mRNAs) and several important classes of non-coding RNAs, such as micro RNAs (miRNAs), long non-coding RNAs (lncRNAs) and small nuclear RNAs (snRNAs). Messenger RNAs undergo post-transcriptional modifications before they are exported to the cytosol where the mRNA is translated into proteins. RNA polymerase III

transcribes the 5S ribosomal RNA, transfer RNA (tRNA) genes and other small nuclear RNAs.

RNAPII is a large multi-subunit enzyme. Its largest subunit RBP1 possesses a carboxy-terminal domain (CTD) which consists of a species-dependent number of heptapeptide repeats: $Tyrosine_1$-$Serine_2$-$Proline_3$-$Threonine_4$-$Serine_5$-$Proline_6$-$Serine_7$ ($Y_1S_2P_3T_4S_5P_6S_7$; Corden et al., 1985). The CTD is not required for the catalytic function of the RNAPII, but it is crucial in regulating the multiple steps of transcription and coupling it with co-transcriptional RNA processing (Hsin & Manley, 2012). Different post-translational modifications of the CTD are associated with different stages of the transcription cycle (Brookes & Pombo, 2009).

Enhancers

Enhancers are DNA sequences that can act in *cis* to regulate transcription of specific genes in a cell-type and/or developmental stage specific manner, irrespective of their orientation or linear distance from their target gene promoters (Beagrie & Pombo, 2016). To achieve their function, enhancers are brought in close proximity to the promoters through the formation of chromatin loops. This process is mediated by one or a combination of transcription factors (TFs), co-factors and chromatin remodelling factors (Banerji et al., 1981; Fraser et al., 2015; Pombo and Dillon, 2015). It is estimated that the human genome contains hundreds of thousands of enhancers, which widely exceeds the estimated 20 thousand protein-encoding genes (Pennacchio et al., 2013). This observation, together with various studies in different cell types suggest that many genes are likely regulated by multiple enhancers, contributing to the fine tuning of cell-type specific gene expression (ENCODE Project Consortium, 2012; Rao et al., 2014).

Chromatin factors and histone modifications that are commonly used to identify putative enhancers include histone acetyl transferase (p300), Histone 3 lysine 27 acetylation (H3K27ac) and Histone 2 lysine 4 monomethylation (H3K4me1) (Pennacchio et al., 2013). Histone modifications can open local chromatin; for example acetylation of H3K27 reduces the affinity between nucleosome components, making the DNA more accessible to TFs binding. Modified histones also act as platforms to which factors involved in enhancer function can bind. Enhancers are bound by lineage-specific TFs and RNAPII (Koch et al.,

2011, Spitz & Furlong, 2012). Binding of multiple TFs can have activating or silencing effects on the target gene promoter. The combinatorial binding of different TFs also specifies the function of the enhancer in time and space (Small et al., 1992).

Regions of chromatin are occupied with different numbers of nucleosomes, which strongly influences their compaction. Enhancers are typically nucleosome depleted, which leaves free DNA that can be targeted by transcription factors. Nucleosome depleted regions can be detected by enzymatic digestions, such as deoxyribonuclease I (DNase I), and are used as marks for accessible chromatin and enhancers (Pennacchio et al., 2013).

Gene activation by enhancers

The precise mechanism of how enhancers modulate gene expression remains to be discovered, but different models have been proposed. Here, I am going to discuss three of these mechanisms. In the transactivation model, the transcription pre-initiation complex assembles at the gene promoter followed by RNAPII binding. In this context, transcription of the gene is stimulated upon direct interaction with the TFs bound to the enhancer. The exact mechanism of how the effector domain of the TF acts on the transcription complex is not known, but various models have been proposed. For example, it can stabilize the pre-initiation complex or increase the rate of transcription elongation. Another mechanism proposed is the hit-and-run. In this model the interaction between the enhancer and promoter results in deposition of chromatin remodelling enzymes on the promoter. These enzymes can induce nucleosome remodelling through deposition of activating histone marks, and ultimately result in an open chromatin state. In this model enhancer-promoter loops are necessary for activating the promoter, but not for progressive transcription to occur (both reviewed in Beagrie & Pombo, 2016). An alternative model that does not imply a direct enhancer promoter contact is the polymerase tracking model, where RNAPII binds to the enhancer and moves along the DNA until it reaches the target gene promoter. Once the enhancer encounters the promoter, the RNAPII starts transcribing the gene (reviewed in Beagrie & Pombo, 2016). This mechanism is likely to occur only between enhancers and promoters that are located in physical distance, and it appears to be less transcriptionally efficient, as genes engaged in promoter-enhancer loops show higher levels of expression when compared to genes lacking promoter-enhancer loops (Rao et al., 2014).

Irrespective of the mechanism through which enhancers modulate transcription, several

factors have been identified to play a role in this process (Figure 1.1). For example, an important coactivator is the multiprotein complex mediator, which can act as a signal integrator for activating enhancers (Allen et al., 2015). The mediator complex has been found to be present and important for efficient transcription of many gene-coding regions. Presence of mediator is required for activation of transcription, where mediator acts by helping to form the pre-initiation complex, stimulating phosphorylation of the RNAPII CTD and promoting elongation (Conaway & Conaway, 2011). Mediator can also physically bridge the enhancers bound by TFs with the promoters bound by the transcription machinery (Allen et al., 2015). Another important co-factor for enhancer-promoter activation is cohesin. Cohesin is brought to active enhancers by the cohesin-loading factor NIPBL, and forms a ring complex around the enhancer-promoter loops, transiently stabilizing the interaction (Weintraub et al., 2017).

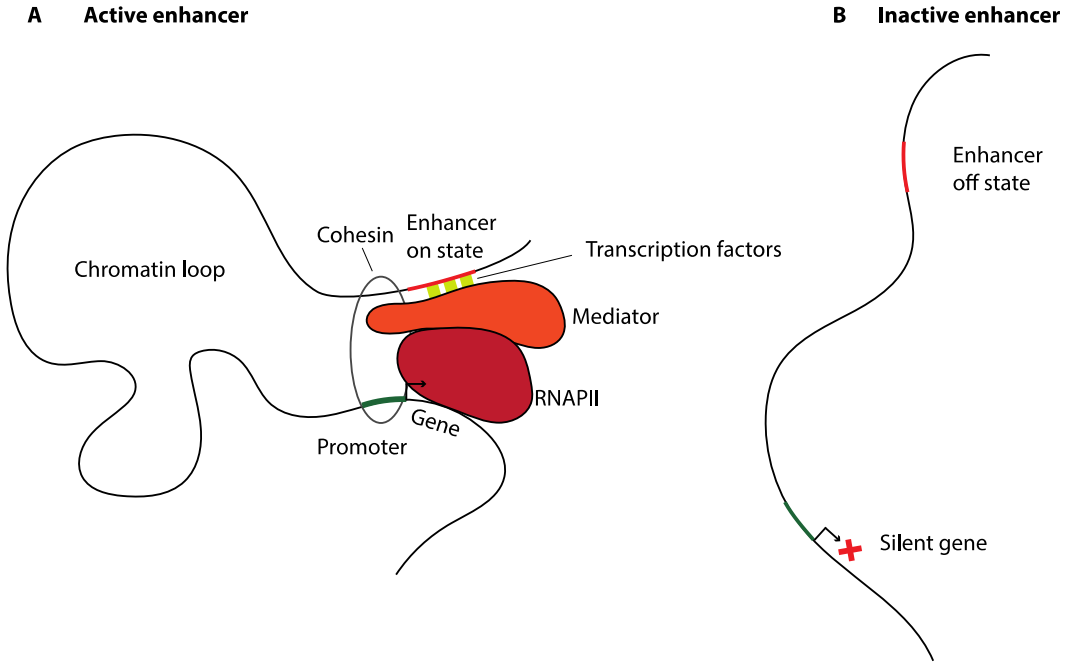


Figure 1.1: Gene activation at enhancers
A Schematic representation of an enhancer-promoter loop. An active enhancer contacts a promoter by looping interaction. The mediator complex facilitates bridging between the RNAPII bound promoter and the active enhancer bound by transcription factors. Cohesin forms a ring around the loop, stabilizing the interaction. This complex facilitates activation and maintenance of transcription.
B Schematic representation of a gene and its enhancer in an inactive form. In the inactive enhancer state, no loop is formed between the enhancer and the promoter, hence the gene is not transcribed.

Locus control regions and super-enhancers

The locus control region (LCR) is a stretch of *cis* regulatory DNA that modulates the expression of genes that can be located further away. The LCR shows a tissue- and developmental stage-specific function. LCRs drive expression of reporter genes independent of their location, and have been detected in many species and tissues, often with conserved sequences. One of the best studied examples is the LCR that regulates the expression of the β-globin genes in the erythroid cells. The expression of the globin genes at the human β-globin locus occurs in a developmentally regulated manner, and is controlled by five highly conserved enhancer elements, clustered in the direct genomic vicinity. Globin gene expression is detected when the enhancers are contacting the gene promoters (Li et al., 2002; Levings et al., 2002).

Recent work has identified a new class of enhancers, so called super-enhancers which are defined in embryonic stem cells (ESCs) as regions with unusually high transcription factor and mediator binding, as well as occupying larger genomic regions than traditional enhancers, the latter feature also giving them the name stretch enhancers. Super-enhancers are very strong enhancers that are thought to regulate cell-type specific genes and have been identified as contributing in several diseases (Whyte et al., 2013; Hnisz et al., 2013, Loven et al., 2013, Hnisz et al., 2015, Hamdan & Johnsen 2017). In stem cells for example, key pluripotency genes such as Oct4 and Nanog are regulated by super-enhancers. The BET protein BRD4 occupies the super-enhancers of these core stem cell genes and regulates their expression, by recruiting mediator and CDK9 (Di Miccio et al., 2014).

Super-enhancers were first defined in mouse ESCs in Whyte et al. (2013) as regions containing: (1) sites bound by all three master regulators Oct4, Sox2 and Nanog; (2) enhancers within 12.5 kb were stitched to define a single entity spanning a genomic region; (3) stitched enhancers and the remaining individual enhancers were ranked by the background level of Med1 signal within the genomic region. Those above the Med1 cut-off level were considered super-enhancers, those below the cut-off as normal enhancers (Whyte et al., 2013). Besides Med1 increased levels of RNAPII, CBP, p300, cohesin and several activating histone marks such as H3K27ac, H3K4me2 and increased chromatin accessibility as measured by DNase-seq were detected at super-enhancers (Hnisz et al., 2013). Super-enhancers bear resemblance to LCRs and it is under debate whether they describe the same

entity (Pott & Lieb, 2014). It is also not yet fully understood to what extent super-enhancers are different from typical enhancers, and whether the individual enhancers constituents within the super-enhancers cluster function synergistically. Moorthy et al. reported that deleting complete and partial super-enhancer clusters, leads to highly variable reduction in their target gene expression, ranging from 12% to 92% (Moorthy et al., 2016). Furthermore, deletion of individual enhancers from the super-enhancers or deletion of sub-clusters revealed partial redundancy of super-enhancer components. A similar result has been reported by Hay et al., where functional dissection of individual enhancers at the α-globin super-enhancer cluster showed that individual enhancers contributed differentially to the α-globin expression (Hay et al., 2016). A possible explanation for the presence of multiple enhancers could be an added contribution in providing robustness in gene expression (Hay et al., 2016).
Super-enhancers have been detected in different cell types, where they have been shown to regulate key cell identity genes. Gene ontology analysis conducted on sets of genes associated with super-enhancers in various cell types such as ESCs, myotube, macrophage, pro-B cells and T helper cells, shows that in each cell type, the top ten most significant biological process terms were crucial for the specific function of each cell type (Hnisz et al., 2013).

1.3 Nuclear organization and folding of the chromatin fibre in mammalian cells

How are the approx. 2 meters of DNA fibre packed into a cell nucleus of only several micrometres in diameter? To fit into the nucleus, the DNA needs to be condensed by a factor of several hundred thousand (>300 000 fold for humans). This is achieved by folding the string of DNA into subsequently higher order-structures, that increase its compaction. At the smallest scale, the DNA is wrapped around nucleosomes. The DNA-protein complex called chromatin is further folded by interactions between enhancers and promoters, which are often within highly self-interacting genomic regions called topologically associating domains (TADs). TADs have been described as a basic unit of genome organization (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). TADs contact other TADs in a preferential way forming metaTADs (Fraser et al., 2015). At a larger scale, the genome is further organized into compartments. These are large megabase-sized genomic regions that have been found to differ in their basic epigenetic characteristics as being transcriptionally active (A-type) or transcriptionally silent (B-type). Regions that are alike in their compartment-type engage in

long-range interactions (Liebermann-Aiden et al., 2009). To achieve a maximal degree of condensation, it has been suggested that DNA supercoils into a compact fibre, which ultimately forms the basis of the folded chromosome (Gilbert & Allan, 2014).

The many levels of organization that chromatin is engaged in are linked with different functions and provide environments that prime the chromatin in a specific way. For example, gene expression and silencing are associated with different spatial organizations. Actively expressed genes often engage in looping interactions with enhancer elements. To provide sufficient space for these factors to bind, transcriptionally active regions need to be more open and decondensed. This is supported by activating histone marks, such as Histone 3 lysine 27 acetylation (H3K27ac), Histone 3 lysine 4 monomethylation (H3K4me1). Silent heterochromatin is highly compacted, leading to steric exclusion of RNA polymerase II and leaving the regions transcriptionally silent. Heterochromatin is marked by Histone 3 lysine 9 (H3K9) methylation, heterochromatin protein 1 (HP1) and is enriched in repetitive sequences. Heterochromatin is an important constituent of the genome, with roles in genome architecture and stability. Additional plasticity and functional clustering of activity is achieved by the association of individual genes or large stretches of DNA with nuclear landmarks, such as transcription factories and the nuclear lamina (Bickmore & van Steensel, 2013; Pombo and Dillon, 2015). A general picture emerges, where the position of a gene within the nucleus, together with its genomic environment shape its activity. Many examples of dynamic relocation of genes after gene activation or silencing have been reported. For example repositioning of the activated gene towards the nuclear interior for IgH in B lymphocytes (Kosak et al., 2002) and Mash1 in neuronal cells (Williams et al., 2006) or relocation to the nuclear periphery upon silencing (Peric-Hupkes et al., 2010). In conclusion, bringing the gene into a different nuclear environment, as well as changing its environment (e.g. by changes in compaction, rewiring of genomic contacts and/or nuclear relocation) can have transcriptional implications on the gene, and ultimately cellular fate.

Topologically Associating Domains

TADs are described as genomic regions that contain DNA sequences which physically interact with each other more often within the TAD than with sequences outside of the TAD. They were discovered by 3C-based methods, which determine interaction frequencies

between genomic regions (Nora et al. 2012; Dixon et al., 2012). TADs appear as separate domains of high internal interaction frequency and an average length of ~ 500kb – 1 Mb (Nora et al., 2012; Dixon et al., 2012). FISH experiments confirmed, that a set of two DNA probes covering the interior of a TAD colocalise more often than two probes spanning a set of neighbouring TADs with a boundary in between (Figure 1.2; Nora et al., 2012).

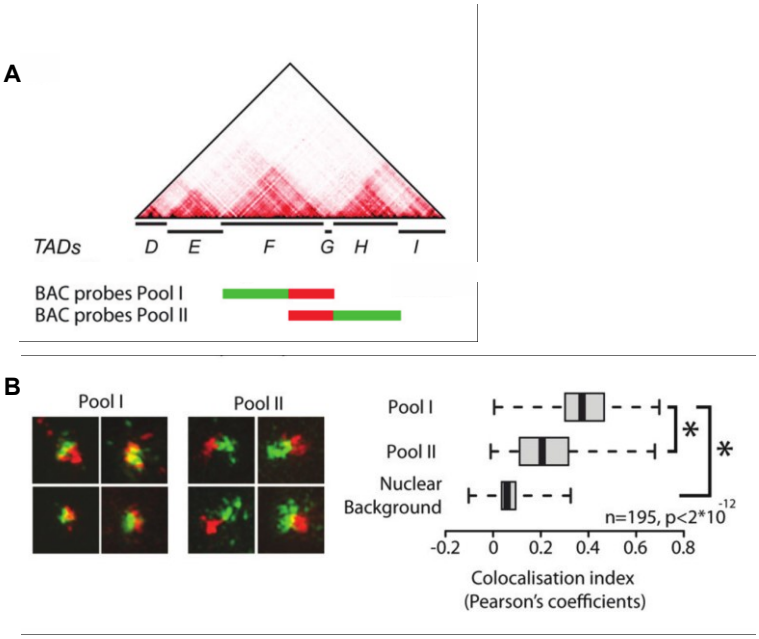


Figure 1.2: FISH confirms the spatial separation of TADs
A Positions of FISH probes in relation to the underlying TADs. BAC probes pool I cover TAD F and the small TAD G. BAC probes pool II cover the boundary between TADs H, G and F.
B Structured illumination microscopy revealed that colocalisation of neighbouring sequences was greater when they belong to the same TAD F (and G). Boxplots display the distribution of Pearson correlation coefficient between red and green channels, with whiskers and boxes encompassing all and 50% of values respectively; central bars denote the median correlation coefficient.
(modified from Nora et al 2012)

TAD boundaries are defined by the directionality index (DI), a measure of the degree of upstream and downstream interaction bias (Dixon et al., 2012). They are enriched for many elements such as the DNA insulator protein CTCF, Cohesin, SINE elements, housekeeping genes, tRNAs, retrotransposons, RNAPII, different histone marks such as Histone 3 lysine 36 trimethylation (H3K36me3), and different transcription factors (Dixon et al., 2012). Amongst these, CTCF and Cohesin have been suggested to have important functional roles in TAD formation. CTCF is an 11-zinc finger DNA binding protein with a central zinc-finger domain. Cohesin is a ring shaped multi-protein complex composed of Smc1, Smc3, Rad21 and SA1/2.

Many TADs display peaks in interaction frequency correlating with their boundaries, suggesting a strong contact at the TAD boundaries. This observation has been linked with a model where CTCF binds to its specific sites, followed by cohesin and resulting in the formation of a loop containing the TAD (Rao et al., 2014). The central role Cohesin plays in sustaining TADs was recently shown; deletion of the Cohesin loading complex Nipbl leads to loss of TAD structures (Schwarzer et al., 2017). During the mammalian cell cycle, TADs form and disassemble, highlighting that over time TADs are dynamic structures (Nagano et al., 2017).

TADs are one of the most basic DNA-folding units and widely utilized to describe local folding as well as whole chromosome organization. Despite the growing knowledge about TADs, many fundamental aspects about TAD organization are still unknown, such as, at what physical distances different parts of a TAD localize, how restrictive TAD boundaries are with respect to interactions amongst neighbouring TADs and what do TADs represent at the single cell level.

SubTADs

TADs are thought to contain local interactions between enhancers and promoters. These interactions are often contained within domains, at the sub-megabase scale (subTADs) (Phillips-Cremins et al., 2013; Berlivet et al., 2013; Dowen et al., 2014). Functional clustering of enhancers and their target genes in specific cell types occurs in subTADs. In different cell types, the subTAD topology is altered, representing modified gene-enhancer and enhancer-enhancer interactions. SubTAD interactions might be at the base of gene regulatory networks that coordinate expression. Cell type specific subTAD formation has been shown for example during limb development, where expression of HoxA genes is controlled by multiple enhancers located upstream of the HoxA gene cluster. In limb cells, these enhancers are grouped into distinct subTADs where they physically interact with their HoxA target genes (Berlivet et al., 2013). The structure of the subTAD is anchored by CTCF-mediated looping interactions (Dowen et al., 2014). SubTADs are thought to create isolated neighbourhoods that may restrict enhancer-promoter contacts. Disruption of local domain boundaries has been shown to induce enhancer looping and ectopic gene activation, outside the subTAD/TAD neighbourhood (Dowen et al., 2014; Lupiáñez et al., 2015; Flavahan et al., 2016; Hnisz et al., 2016).

MetaTADs

Interactions between TADs also known and inter-TAD interactions can be presented in a hierarchical way, where each TAD has preferred interaction partners. This preferential domains-within-domains hierarchy is called a metaTAD tree (Figure 1.3). In Figure 1.3A, a region on chromosome 2 is shown that is divided into the underlying TADs (1-6) based on the directionality index (DI). The preferential contacts between the neighbouring TADs can be represented as a tree like structure (Figure 1.3B). Increasing tree levels correlate with an increased folding of preferred TAD interaction partners into larger domains, thereby achieving efficient chromatin packaging.

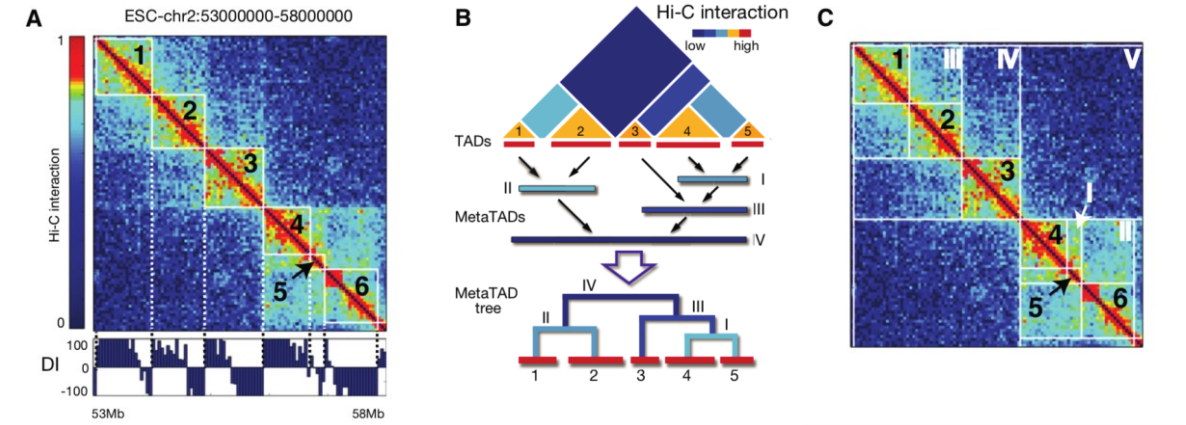


Figure 1.3: TADs engage in preferred inter-TAD interactions
A Hi-C map of a region on chromosome 2 in mESCs. The directionality index (DI) defines the TADs 1-6.
B Schematic representation of a Hi-C interaction matrix (top) showing metaTADs and metaTAD tree arising from the Hi-C interactions (bottom). The preferential interactions of TADs are represented as branches in the tree. Preferential interactions between branches build up subsequent tree levels.
C Examples of TADs (1–6) and metaTADs (I–V) in the same region shown in (A).
(adapted from Fraser J. et al 2016)

The tree structure also correlates with genetic, epigenetic and expression features, such as replication timing, lamina association, histone marks associated with active transcription (e.g. H3K4me3, H3K27ac, H3K36me3), CTCF and several transcription factors. TADs within a metaTAD also frequently belong to the same A/B compartment, especially within the lower tree levels (Fraser et al., 2015). MetaTAD topology can change during the neuronal differentiation, where about 20% of the TADs undergo structural changes (Figure 1.4). Structural tree rearrangements during differentiation are linked to transcription changes, showing a link between long-range chromosome folding and gene regulation (Fraser et al., 2015).

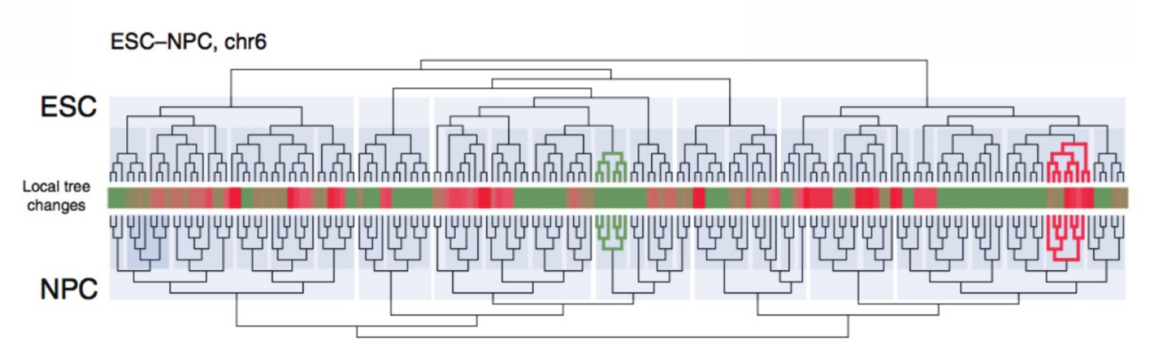


Figure 1.4: The metaTAD trees of mouse ESCs and neurons for chromosome 11
Comparison between the metaTAD trees of chromosome 6 in ESC and neuronal progenitor cells. The extent of local tree topology changes is represented in the central heatmap, where red indicates tree changes and green tree conservation. Some examples for tree conservation (green) and changes in the tree wiring (red) are shown.
(adapted from Fraser et al., 2015)

Compartments

An important general principle of chromosome folding at the megabase level is the formation of two types of compartments: A- and B-type compartments. Chromatin shows a strong preference for contacts between regions of the same compartment type. Identification of loci within each compartment type was performed by principle components analysis (PCA) (Lieberman-Aiden et al., 2009). Compartment A consists of open chromatin, characterized by active histone marks. It is gene-rich, with high GC content, susceptible to DNase hypersensitivity and typically consists of regions that are early replicating. Conversely, compartment B, consists of gene-poor, late-replicating genomic regions, covered with gene-silencing histone marks and a more compact configuration. For example, heterochromatin and lamina-associated domains (LADs) are typically found in the B-type compartment. Compartment changes that occur during differentiation are linked to transcriptional changes within the same genomic regions (Dixon et al., 2013). High resolution Hi-C studies further divided A/B compartments into five sub-compartments: two of A-type and three of B-type (Rao et al., 2014). At large genomic distances and at the chromosome scale, the folding of TADs of A/B character is organized in a spatially polar manner, where compartments A and B cluster in separate domains (Wang et al., 2016).

The formation of A/B compartments is independent of TAD formation, as loss of TAD structure by deleting the cohesin loading factor Nipbl, preserves and even reinforces

compartments (Schwarzer et al., 2017). TADs and PCA compartments seem to therefore constitute two independent, fundamental ways the genome organizes.

Chromosome territories

Chromosomes have preferred positions within the cell nucleus, which are called chromosome territories (Figure 1.5). Chromosome territories are highly conserved within the same cell type amongst different species, implying a fundamental role in the functioning of the genome (Tanabe et al., 2002). Although it is unclear, how chromosome territories are formed**,** as a general principle gene dense, transcriptionally active and early replicating regions tend to cluster in the nuclear interior, whereas more silent and gene poor chromosomes locate at the nuclear periphery (Croft et al., 1999; Boyle et al., 2001; Grasser et al., 2008). Exception to this general organizational principle are rod/cone cells in nocturnal animals. In these cell types an inverted nuclear organization occurs, where the euchromatin locates at the nuclear periphery and the heterochromatin is positioned in the nuclear interior (Solovei et al., 2009). The inverted organization serves the physical purpose of bundling the incoming photons, by using the densely packed chromatin as a lens that aids to focus the light shining through the rod cells.

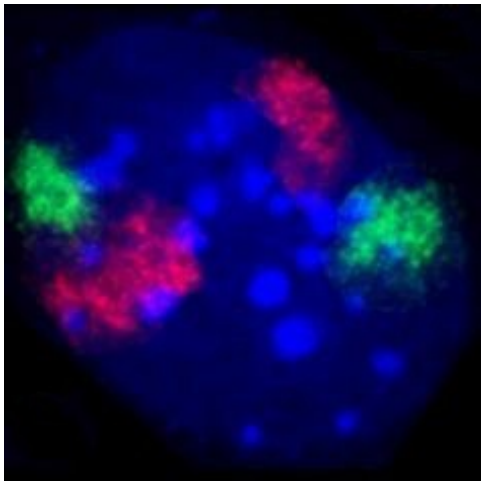


Figure 1.5: Chromosome territories
Chromosome territories are distinct, non-random positions that chromosomes occupy in the nucleus. The image shows mouse fibroblasts, the nuclear DAPI staining is shown in blue, chromosome 2 is shown in red and chromosome in 9 green. (Image taken from Mayer et al., 2005)

Chromosomes can intermingle at their peripheries. In the inter-chromatin network model (ICN), loops from neighbouring chromosomes can reach out of their territories and extend into the neighbouring chromosome territory (Branco & Pombo, 2006). This way active genes could loop out and colocalize in *trans* with active genes of neighbouring chromosomes for co-regulation, as observed in the case of transcription factories (Branco & Pombo 2006).

Some particularly active gene loci, such as the Hox genes, have been found to reside at the periphery of the territory and they loop out from the core territory upon activation (Chambeyron et al., 2005).

Models to describe Chromosome folding

One model to describe the folding of whole chromosomes is the fractal globule model, which successfully describes certain characteristics of chromatin folding and allows to determine ensembles of possible conformations (Lieberman-Aiden et al., 2009). The model originates from polymer physics, where it describes the spontaneous folding of a polymer as a result of topological constraints, preventing one region of the chain from passing across another. The fractal globule model is described by a hierarchy of crumples forming a self-similar structure and folding follows the power law (Grosberg et al., 1988). A study in Drosophila has shown that the fractal globule model can be used to describe the folding of regions with different epigenetic states, where each epigenetic state has a distinct power-law coefficient. Polycomb repressed domains have a scaling coefficient smaller than one, indicating an increase in compaction with increasing genomic length. The scaling coefficient of active domains is greater than one, indicating a decrease in compaction with increasing genomic length. For inactive domains, the scaling coefficient is equal to one, hence the 3D density over different lengths remains constant (Boettiger et al., 2016).

At the TAD level, which is at a length scale of 500 kb to 1 Mb, a popular model to describe chromatin folding is the loop extrusion model. This is based on bringing together CTCF boundary elements by extracting the intervening DNA with a motor protein, thereby creating a large loop of chromatin (Rao et al., 2014).

An alternative model for local folding is the strings and binders switch model, which is based on interactions between specific regions along the DNA fibre following basic thermodynamic principles (Barbieri et al., 2013). In Barbieri et al. (2017), the strings and binders switch model was used to simulate different scenarios of chromatin contacts at the HoxB locus. In this model, contacts were driven by: i) active or polycomb-repressed promoter states, ii) only the presence of active RNAPII post-translational modifications (RNAPII-S5p and/or H3K4me3, without the contribution of polycomb), and iii) CTCF occupancy (Barbieri et al., 2017). Results suggested that active and polycomb-repressed promoter states play separate

but simultaneous roles in the folding of the HoxB locus at the genomic length scale of tens of kilobases. CTCF occupancy alone is not sufficient to reconstitute locus folding at this scale. It has been shown that CTCF plays an important role in loop formation, its contribution to the long-range folding of chromatin and TAD formation however remains under debate.

1.4 Structural components of nuclear organization

The nucleus consists of a complex non-homogenous environment and different sub-compartments are associated with different functions and activity states of genes positioned within these compartments. Nuclear landmarks are functional regions that cluster specific activities, providing a scaffold to organize chromatin and induce a greater efficiency in the functional output that is required. Nuclear landmarks are therefore important in contributing to genome organization on many genomic length-scales.

Transcription factories

Transcription by RNAPII occurs in the nucleoplasm at discrete sites called transcription factories (Iborra et al., 1996). They are ~ 50 nm in size and contain on average clusters of five to eight RNAPII molecules per factory (Martin & Pombo, 2003). Transcription factories create hubs of transcription, where active transcription units are co-transcribed (Noordermeer et al., 2008; Ferrai et al., 2010). A single transcription factory could thereby transcribe several genes at the same time in *cis,* and in *trans*, which could explain why the number genes transcribed at any moment in time is considerably larger than the number of transcription factories. Characterization of the spatiotemporal RNAPII organization showed that transcription factories form transiently, with an average lifetime of 5.1 (± 0.4) seconds. Upon stimuli that induce transcription, an increase in the number of RNAPII clusters was detected (Cisse et al., 2013).

Transcription factories are thought to have a proteinaceous core, with the RNAPII molecules facing outside. Upon isolation, transcription factories were found to contain transcription regulators such as co-activators and chromatin remodellers, TFs, histone modification enzymes, ribonucleoproteins, as well as splicing and processing factors (Melnik et al., 2011).

Splicing speckles

Splicing speckles are nuclear compartments, containing pre-mRNA splicing factors, such as SC35, ASF1/SF2 and the splicing co-activator SRm300. In total, there are approximately 20-30 speckles per nucleus. Typically, speckles reside between the chromosome territories, rendering them also the name inter-chromatin granules (Hall et al., 2006). Spatial distribution of splicing speckles can be studied by immunofluorescence using antibodies specific for SC-35 and Sm splicing factors (e.g. Xie et al., 2006).

Speckles are enriched in poly(A) RNA, which comprises mostly mRNA, suggesting that mRNA metabolism is associated with these domains. Multiple different mRNA transcripts originating from the same or different chromosomes have been found within the same speckle (Hall et al*.,* 2006). Furthermore, genes from the same or different chromosomes were found to colocalize at the same speckle periphery (Shopland et al., 2013). The association of genes with speckles is not simply a consequence of a need of splicing factors for proper splicing, but is specific, as not all genes associate with the speckle (Hall et al., 2006). Most introns of genes that are found at speckles are removed by co-transcriptional splicing, before the mRNA proceeds into the centre of the speckle, however one or a few introns can still be present. It has been suggested that speckles are quality control checkpoints, where incorrectly spliced mRNA is retained and correctly spliced mRNA is prepared for nuclear export (Hall et al., 2006). RNAPII has been found within speckles, however its role has not yet been understood, as the speckle interior itself is transcriptionally inactive (Puvion & Puvion-Dutilleul 1996; Cmarko et al., 1999; Xie et al., 2006). Presence of RNAPII may however serve to couple splicing at the speckle with transcription.

Nuclear lamina

The nuclear lamina is a repressive nuclear compartment and genes that associate with the lamina are typically transcriptionally silent (Guelen et al*.,* 2008). LADs, tend to be late replicating and are usually part of the B-type compartments. The nuclear lamina covers the nucleoplasmic side of the inner nuclear membrane and consists of long proteinaceous polymers (Prokocimer et al., 2009). The genomic regions contacting the lamina, the LADs, were mapped using DNA adenine methyltransferase identification (DamID) (Guelen et al., 2008). LADs are cell type specific, they have a median size of ~0.5 Mb and cover in total about 40 % of the genome (Bickmore & van Steensel 2016).

Nucleolus

Transcription of rRNA by RNAPI, subsequent rRNA maturation and ribosome production, occurs in the nucleolus. The nucleolus has a distinct ultrastructure, where rRNA transcription and processing occur in dedicated compartments. The three compartments are: the fibrillar centre (FC), the dense fibrillar component (DFC) and the granular component (GC). The FC contains RNAPI and its associated co-factors. The template rRNA is positioned at the interface between the FC and the DFC, as nascent rRNA transcripts migrate into the DFC, where they are processed. The GC is the location of ribosome maturation (Martin & Pombo 2003; McKeown & Shaw 2009). Interestingly, not only rDNA localizes at the nucleolus, but also stretches of genomic DNA containing repressed genes, marked by the repressive histone mark H3K9me3 (Nemeth et al., 2010; van Konigsbruggen et al., 2010). In accordance with LADs, those regions are termed nucleolus associated domains (NADs). NADs and LADs are most likely consisting of a similar repressive chromatin type.

1.5 Genomic rearrangements

Genomes are structurally varied and variations of intermediate size, ranging from kb to Mb scale, including deletions, duplications and inversions of DNA segments (Figure 1.6), as well as translocations, have been recognized to contribute substantially to natural human genetic variations (Sebat et al., 2004). In particular, genomic variability due to deletions and duplications has been found to account for 5 to 10 % of the overall variability in the human genome (Zarrei et al., 2015). Structural variations (SVs) have also been linked to an ever-rising number of human genetic disorders, including intellectual disabilities, congenital malformations and cancer (Campbell et al., 2008; Kurth et al., 2009; Craddock et al., 2010). SVs are not only reserved to the coding portion of the genome, but also non-coding regions have been connected to human disorders (Klopocki & Mundlos 2011).

Genomic rearrangements and the reorganization of DNA in the context of disease are of great interest in the field of chromatin organization, as SVs can interfere with gene regulation and chromatin architecture. Phenotypes arising from SVs show the implications of disrupting the local chromatin architecture, leading to the reconfiguration of 3D chromatin, which can bring regulatory elements into the proximity of genes and triggering their ectopic expression (Lupianez et al., 2015; Franke et al., 2016). Conversely, duplications can span large regions but not have any phenotypic implications on the organism (Franke et al., 2016).

Understanding the underlying rewiring of the genome, in particular how the genome organization changes, e.g. upon insertion of a duplication, can explain why some SVs lead to abnormal phenotypes and some do not. Furthermore, duplications raise the question how does the local chromatin environment rearrange to accommodate the additional DNA, and whether the duplicated region simply integrates without disrupting previous interactions. It is also still unknown whether duplications lead to a homogenous new folding state of the DNA or whether an ensemble of structures arises from these SVs.

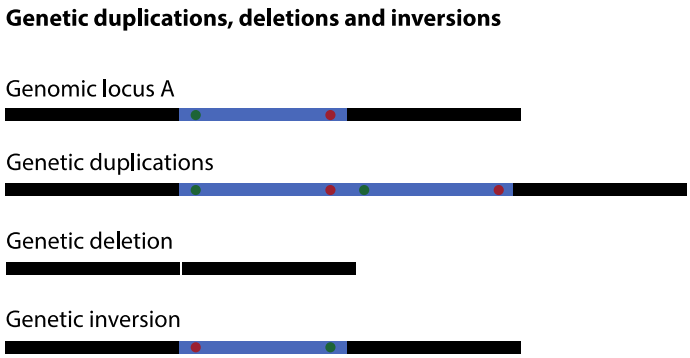


Figure 1. 6: Genomic rearrangements alter the genomic sequence
A multitude of genomic rearrangements exist which introduce structural changes in the genome, that can lead to disease. Genomic duplications lead to doubling of the genetic content. Deletions remove genetic content and inversions rewire the underpinned region by inverting the DNA. Structural variations can bring genes and enhancers into a new chromatin environment and thereby potentially initiate ectopic gene expression.

An important developmental locus that contains several known genomic rearrangements is the Sox9 locus. Sox9 is crucial for male sex development and chondrogenesis. The TAD harbouring the Sox9 gene also contains a large gene desert with multiple regulatory elements and sites associated with human diseases (Gordon et al., 2009). Duplications including the RevSex region, which is located upstream of Sox9, induce male-to-female sex reversal. A second duplication which includes the RevSex region, but extends into the neighbouring TAD overlapping with the Kcnj 2 and Kcnj 8 genes, results in a limb malformation phenotype, known as Cooks syndrome (Kurth et al., 2009). Interestingly, a duplication without a phenotype is also known at the Sox9 locus. It contains the RevSex region, the gene desert and stretches into the neighbouring TAD but does not include the Kcnj genes. Carriers of this structural variation are phenotypically normal even though the duplication overlaps in

large parts with the disease implicated duplications. Understanding how the chromatin rewires as a result of a duplication and how this influences ectopic gene expression is an important quest in solving the genomic disease mechanisms.

1.6 Methods to investigate chromatin architecture

Several methods to investigate genome architecture have been developed in the past decade, especially since the recent advances in sequencing technologies. Three important methods to study genome architecture are Imaging based approaches, Chromosome Conformation Capture (3C)-based technologies, such as Hi-C, and Genome Architecture Mapping (GAM). A very recent further addition to the collection of techniques is the method SPRITE (Quinodoz et al., 2017 biorxiv). Imaging has been crucial in providing important insights into the 3D genome structure in single cells, especially high-resolution imaging techniques, such as cryoFISH, and super-resolution methods such as STED and STORM. Imaging with FISH measures distances between genomic regions and allows mapping of genome topology at a single cell level. Hi-C measures ligation frequencies of genomic regions, within a cell population. GAM is a recently developed technology that provides information about the co-segregation of genomic loci within thin cryosections of a cell nucleus. SPRITE (Split-Pool Recognition of interactions by Tag Extension) measures genome-wide higher-order DNA interactions within the nucleus by ligation of unique adapter sequences to crosslinked clusters of chromatin contacts (Quinodoz et al., 2017 biorxiv). The different methods that are available for studying genome architecture are in many aspects orthogonal and work together to answer questions of 3D genome architecture folding.

FISH - Fluorescence in situ hybridization

Fluorescence in situ hybridization (FISH) is an imaging-based technique to visualize the location of genomic regions by hybridizing defined genomic regions with a complementary fluorescently labelled probe. It is based on the same principle as all DNA hybridization methods, taking the ability of single-stranded DNA to anneal to complementary DNA. In FISH, the target DNA is found inside whole cells or tissue sections attached to a glass microscope slide. The use of different DNA probes labelled with different fluorophores allows simultaneous detection of several targets.

FISH is a single cell method, where cell specific information is retrieved. Quantitative image analysis allows to extract inter-probe distances, as well as volumes and their location within

the cell nucleus, in particular with respect to nuclear landmarks. Figure 1.7 is a schematic overview of the crucial steps which are part of the FISH protocol.

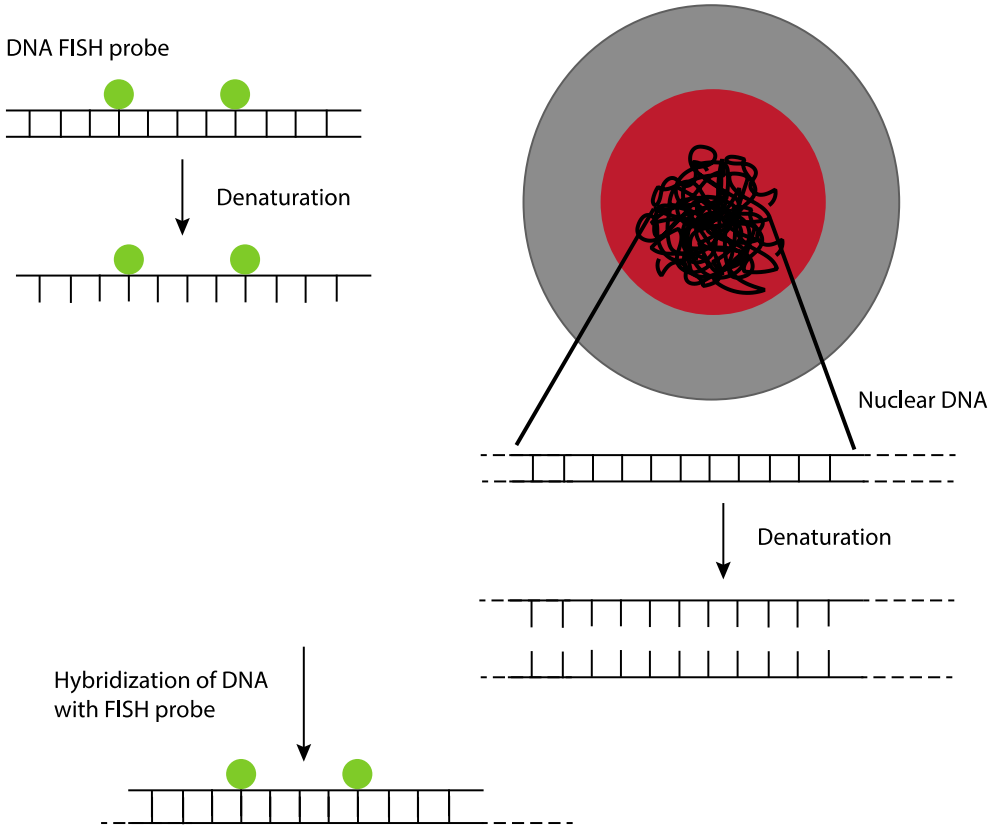


Figure 1.7: Fluorescence in situ hybridization (FISH)
Schematic representation of the FISH technique. The DNA FISH probe is complementary to its target genomic sequence and tagged with a fluorescent marker. Double stranded DNA FISH probes are denatured, to obtain single stranded DNA probes. Nuclear DNA is also denatured to make it accessible to the FISH probes. When the single stranded FISH probe is incubated with the denatured DNA, the probe binds to its complementary DNA sequence and anneals to form a double stranded molecule. Probes can then be imaged at the microscope.

CryoFISH

An addition to the conventional FISH is the high-resolution cryoFISH version, which offers two crucial advantages: an increase in z-resolution and a more stringent fixation than conventional FISH (Branco & Pombo, 2006). Imaging the nucleus and chromosome architecture by fluorescence microscopy requires that the probes can reach into the interior of the nucleus to bind to their target. This in turn puts a limitation/threshold on the degree of fixation that is applicable, since a too strong fixation blocks antibodies and chemical reagents that need to penetrate the nuclear interior. Common solutions to this are a mild fixation or harsh permeabilization, which has been shown to introduce artefacts (Pombo 2007; Guillot et al., 2004; Solovei et al., 2002). Tokuyasu cryosectioning of thin (180 - 200 nm sections) sections, combined with fluorescence microscopy provides increased probe accessibility due to the sectioning, with the additional benefit of a more stringent chemical fixation, which provides a better structural preservation (Guillot et al., 2004; Pombo, 2007). Studies have shown that the cellular ultrastructural fixation in Tokuyasu cryosections compares with that of unfixed material (Griffiths et al., 1983; McDowall et al., 1989). Another major improvement of cryoFISH is that it limits the z-resolution to the thickness of the section. Figure 1.8B shows an example of the improved resolution, by visually comparing the image of a 3D cell (top) with that of a cryosection through the same type of cell. The cryosection shows a greatly improved resolution of the ultrastructure and features such as the nucleolus are clearly visible. Conventional laser scanning confocal microscopes provide at best a resolution of ~200 nm in the *x*- and *y*-axes and ~800 nm in the *z*-axis. In cryosections, the z-resolution equals the section thickness, ~200 nm. In cryoFISH, after fixation, cells are embedded in highly concentrated sucrose, which is a cryoprotectant and then snap frozen in liquid nitrogen.

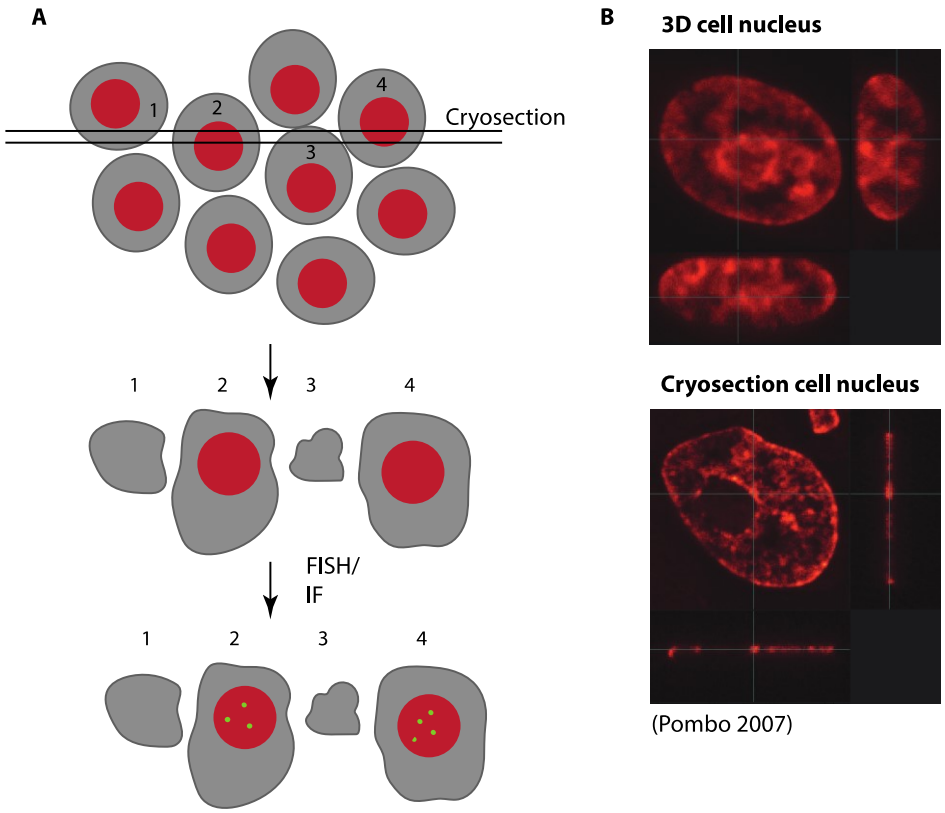


Figure 1.8: CryoFISH Method and resolution improvements
A Cryosections are cut from cryoblocks using a microtome. Sections are collected on a sucrose drop and the sucrose cryo-protectant is removed by subsequent washes with PBS. Cryosections can then be utilized for FISH and immunofluorescence (IF).
B Images of a cell nucleus taken from a 3D sample and cryosection. Major advantaged of cryosections are the improved z-resolution that is limited to the section thickness and the high degree of ultrastructural preservation, due to a more stringent fixation.

Genome architecture mapping

GAM is a recently developed method to study 3D genome organization (Beagrie et al., 2017). GAM is based on the principle that genomic regions which are proximal in 3D will appear more often within the same cryosection, compared to loci that are further apart in the 3D space. Figure 1.9 illustrates this: the proximal regions A and B are often detected within the same section. The more distal locus C however will be less frequently detected in the same section with A or B. Nuclei are randomly selected and cut from cryosections by laser microdissection. This is followed by whole genome amplification and next generation sequencing. The frequency of locus co-segregation, a measure of the frequency of locus co-association within sections, is quantified and pairwise chromatin contacts can be visualized in a heatmap, which is the visual representation of the cumulative detection frequencies of pairs

of genomic regions within the sections (Beagrie et al., 2017). Large numbers of nuclear profiles (NPs) can be collected from different nuclei and contacts can be interred by measuring the frequencies of co-segregation. GAM does not rely on ligation and therefore has the ability to detect the simultaneous co-segregation of multiple genomic regions. Ultrathin cryosections used in GAM are fixed in a way to ensure optimal preservation of the nuclear structure and stabilization of the proteins (Guillot et al., 2004).

To determine non-random contacts that most likely represent specific interactions, the statistical model SLICE was developed. SLICE defines a "probability of interaction" (Pi) as the proportion of a pair of loci, that are considered in an interacting state across a cell population. By calculating the expected co-segregation frequency of any pair of loci the probability of interaction can be background corrected (Beagrie et al., 2017).

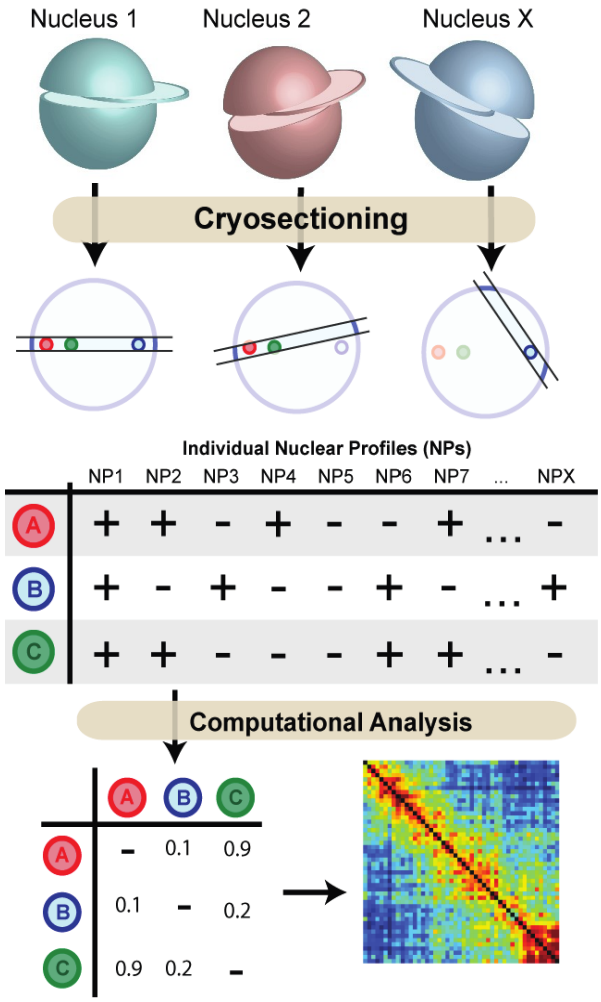


Figure 1.9: GAM method
GAM combines ultrathin cryosections with laser microdissection, high-throughput sequencing and computational analysis. Thin nuclear sections (termed nuclear profiles, NPs) are cut through a population of randomly orientated cells. Loci that are proximal in physical space are more often located within the same section. Using laser microdissection, individual nuclear profiles are collected. After whole genome amplification and sequencing, the genomic content of the individual nuclear profiles is retrieved. Computational analysis of the sequencing data provides information about the frequency of co-segregation of two loci. This frequency can be plotted in a GAM matrix. (Adapted from Beagrie et al., 2017)

Chromosome Conformation Capture

Genome-wide chromosome conformation capture (Hi-C) is a method belonging to the C-technologies, which are population-based methods to study genome architecture. They combine crosslinking of chromatin and proximity ligation of DNA to detect interactions between pairs of genomic loci (Figure 1.10). In Hi-C, the first step involves the fixation of chromatin structure, by crosslinking of DNA with associated proteins using formaldehyde. Subsequently, the crosslinked DNA is digested using restriction enzymes and the ends of the digested DNA fragments are filled using biotinylated nucleotides. Next, the biotinylated DNA is sheared, and DNA fragments containing ligation products are captured by selecting for biotin, before addition of sequencing adaptors to the fragment ends and paired-end sequencing. Thereby, an interaction matrix is built, that reflects the frequency of how often two genomic fragments were ligated in the nucleus (Lieberman-Aiden et al., 2009).

A further development of the Hi-C technique that improves detection of distal sequences that significantly interact is capture Hi-C (cHi-C). cHi-C involves first generating a Hi-C library, biotinylated RNA or DNA oligonucleotide probes are then hybridized to specific sequences of interest within the Hi-C library, followed by affinity purification of the biotinylated probe–library duplexes, washing of bound DNA and finally next generation DNA sequencing (Mifsud al., 2015)

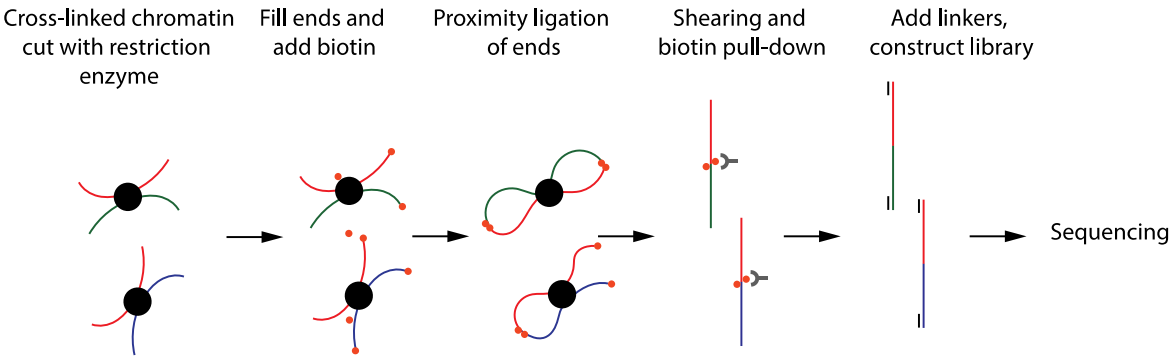


Figure 1.10: The Hi-C method
Cells are cross-linked with formaldehyde, resulting in covalent links between spatially adjacent chromatin segments (DNA fragments: green, red; Proteins, which can mediate such interactions, are shown in black). Chromatin is digested with a restriction enzyme and the resulting sticky ends are filled in with biotinylated nucleotides (red dot). Ligation is performed under extremely dilute conditions to create chimeric molecules. DNA is purified and sheared. Biotinylated junctions are isolated with streptavidin beads and identified by paired-end sequencing.
(adapted from Lieberman-Aiden et al., 2009

2 Materials and Methods

2.1 Cell culture

2.1.1 mESC culturing and neuronal differentiation

Mouse 46C ES cells (a Sox1-GFP derivative of E14tg2a; Ying et al., 2003) were kindly donated by Domingos Henrique (Institute of Molecular Medicine, Lisbon, Portugal). All cryoFISH images were collected from 46C mESCs unless stated otherwise. 46C culture was carried out as previously described (Abranches et al., 2009). Briefly, cells were grown at 37°C in a 5% (v/v) CO 2 incubator in GMEM medium (Invitrogen, Cat# 21710025), supplemented with 10% (v/v) Fetal Calf Serum (FCS; BioScience LifeSciences, Cat# 7.01 batch number 110006), 2,000 U/ml LIF (Millipore, Cat# ESG1107), 0.1 mM beta-mercap- toethanol (Invitrogen, Cat# 31350-010), 2 mM L -glutamine (Invitro- gen, Cat# 25030-024), 1 mM sodium pyruvate (Invitrogen, Cat# 11360039), 1% penicillin-streptomycin (Invitrogen, Cat# 15140122), 1% MEM Non-Essential Amino Acids (Invitrogen, Cat# 11140035) on gelatin-coated (0.1% v/v) Nunc flasks. The medium was changed every day, and cells were split every other day. Before sample collection, ESCs were plated on gelatin-coated (0.1% v/v) Nunc dishes in serum-free ESGRO Complete Clonal Grade Medium (Millipore, Cat# SF001-B) to which was added 1,000 U/ml LIF. ESC batches were tested for mycoplasma infection. mESC culturing and validation was performed by Rieke Kempfer (laboratory of Ana Pombo) and Dorothee Krämer.

2.1.2 Sox9 Duplication cell culturing

The Dup-L mESC cell line was kindly donated by Stefan Mundlos and grown in our laboratory based on protocols provided by the Mundlos laboratory. The growth conditions of ES cells previously generated in the Mundlos laboratory from Dup-L homozygous blastocysts were established in the Pombo laboratory with the help of Rieke Kempfer, using N2B27 Medium supplemented with FGF/Erk, Gsk3 pathway inhibitors (2i) and LIF according to Nagy and Nichols (Nagy & Nichols, 2011). ES cell lines were previously confirmed by PCR genotyping in the Mundlos laboratory.

ES cells were cultured on plates or wells gelatinized and covered with feeder cells, the culturing dishes or wells were coated with 0.1 % gelatin in PBS. After 30 min incubation at 37°C, the gelatin was aspirated and feeder cells were plated at a density of 3-4 x10^4 cells/ cm^2. After at least 6 hours, ES cells were seeded on top of the feeder layer and grown in Knockout Dulbecco's Modified Eagle's Medium (DMEM) 4,500 mg/ml glucose, with sodium pyruvate (Gibco, #10829-018) containing 15 % FCS (PANSera ES, #P30-2600, Lot 130407ES), 1x glutamine (100x, Lonza, #BE17-605E), 1x penicillin/ streptomycin (100x, Lonza, #DE17-603), 1x non-essential amino acids (100x, Gibco, #11140-35), 1x nucleosides (100x, Chemicon, #ES-008D), 0.1 mM β-mercaptoethanol (Gibco, #3150-010) and 1000 U/ ml LIF (Murine Leukemia Inhibitory Factor ESGRO™ (10^7U/ ml, Chemicon, #ESG1107). The ES cell medium was changed every 24 hours and cells were frozen at a density of 1x10^6 cells/ vial in freezing medium, consisting of regular ESC medium supplemented with 20 % FCS and 20 % DMSO (Sigma, #D-2650). Sox9 Duplication culturing and testing for mycoplasma was performed by Rieke Kempfer (laboratory of Ana Pombo). Validation of pluripotency by Immunofluorescence was performed by Dorothee Krämer.

2.1.3 Early neuronal differentiation

The following protocol is taken from Ferrai et al. (2017) and was developed by Carmelo Ferrai (Ana Pombo laboratory). Early neuronal differentiation was optimized based on the method described in Abranches et al. (2009). ESCs were plated with high density (1.5×10^5 cells/cm^2) in serum-free ESGRO Complete Clonal Grade Medium (Millipore, Cat# SF001-B) to which 1,000 U/ml LIF was added. After 24 h, ESCs were washed 3 times with PBS without magnesium and calcium, incubated in PBS for 3 min at room temperature, and then dissociated by incubating in 0.05% (v/v) Trypsin (Gibco, Cat# 25300-054) for 2 min at 37°C. ESCs were plated onto 0.1% (v/v) gelatin-coated 10-cm dishes (Nunc) at 1.6×10^6 cells/dish in RHB-A (Takara-Clontech, Cat# Y40001), changing media every day.

Mouse EpiSCs were established from ESC-46C after growth (4 weeks) in N2B27 basal medium containing 20 ng/ml of Activin (R&D, Cat# 338-AC-050) and 12 ng/ml FGF2 (Peprotech, Cat# 100-18B). The composition of the N2B27 basal medium was as follows: half of DMEM/F12 (Invitrogen, Cat# 21331-020), half of Neurobasal Medium (Invitrogen, Cat# 21103-049), 0.5× N2 (Invitrogen, Cat# 17502-048), 0.5× B27 (Invitrogen, Cat# 12587-010), 0.05 M β-mercaptoethanol (Invitrogen, Cat# 31350-010), and 2 mM L-glutamine (Invitrogen, Cat# 25030-024). EpiSCs were grown on Nunc plates coated with FCS

(BioScience LifeSciences, Cat# 7.01 batch number 110006). Culture medium was changed every day and cells were split every other day, by washing 3 times with PBS without magnesium and calcium, incubating in PBS for 3 min at room temperature, gently scraping them from the plate pipetting up and down 3 times, before transferring to a new FCS-coated plate. The mouse EpiSC-Pitx3-GFP line was previously established in the Li laboratory (Jaeger et al, 2011), from ESC-Pitx3-GFP (Zhao et al, 2004). Frozen cell batches were tested for mycoplasma infection.

EpiSCs were differentiated into day 16 and 30 neurons with midbrain properties using a protocol optimized based on the method developed in Jaeger et al (2011).

The day before starting the differentiation protocol (day "−1"), growing EpiSCs were plated on Nunc plates coated with 15 μg/ml human plasma fibronectin (Millipore, Cat# FC010) and cultured in N2B27 basal medium containing Activin and FGF2, to reach 70–80% confluency after 24 h. Differentiation was started by rinsing cells twice with PBS, and culturing in N2B27 basal medium plus 1 μM PD 0325901 (Axon, Cat# 1408) for 2 days. Medium was refreshed every day. On day 2, cells were washed with PBS, scraped, replated on Nunc 10-cm dishes coated with 15 μg/ml human plasma fibronectin (Millipore, Cat# FC010), and cultured in N2B27 basal medium for 3 days. At this stage, medium was refreshed every day by removing half of the volume and adding half of the volume of freshly prepared medium.

After 72 h, the medium was replaced with N2B27 basal medium plus 100 ng/ml FGF8 (Peprotech, Cat# 100-25-25) and 200 ng/ml Shh (R&D, Cat# 464-sh-025). Medium was refreshed every day by removing half of the volume and adding half of the volume of freshly prepared medium.

After 96 h, cells were washed with PBS, and the medium replaced with N2B27 basal medium plus 10 ng/ml BDNF (R&D, Cat# 450-02-10), 10 ng/ml GDNF (R&D, Cat# 450-10-10), and 200 μM l-ascorbic acid (Sigma, Cat# A4544). Medium was refreshed every day by removing half of the volume and adding half of the volume of freshly prepared medium until days 16 and 30. Cell culturing and validations were performed by Sheila Xie and Carmelo Ferrai (laboratory of Ana Pombo).

2.2 On the use of cryosections

The study of nuclear organisation requires methods of visualisation that are sensitive enough to detect the molecules and genomic regions of interest, and imaging methods that achieve a sufficiently high spatial resolution to allow for subcellular localisation.

Using sectioning of frozen fixed cells, such as Tokuyasu cryosections (Tokuyasu, 1973, 1980), improves the axial resolution on the confocal light microscope from >500 nm to the section thickness. Using cryosections also minimizes chromatic aberration (Pombo et al., 1999). Cryosections furthermore improve probe accessibility (Branco et al., 2006) and provide optimal ultrastructural preservation (Guillot et al., 2004).

2.3 Preparation of cryoblocks

For the preparation of cell blocks for cryosectioning, cells were fixed in 4% and then 8% freshly depolymerised EM-grade paraformaldehyde (PFA) in 250 mM HEPES pH 7.8 (Gibco Cat#15630-056) (10 min and 2 h, respectively). Cell pellets were embedded in 2.1 M sucrose (Sigma-Aldrich® Cat#S9378) in phosphate-buffered saline (PBS) (Sigma-Aldrich® Cat#P4417) for 2 h, before transferring them onto a small metal rod, designed to fit into the holder of the cryosectioning knife. The cell pellet atop the metal rod is frozen in liquid nitrogen and kept in a liquid nitrogen tank (Branco et al., 2006).

2.4 Preparation of cryosections

Cryosections (180-200 nm in thickness, deduced from interference colour) were cut using the Leica EM UC7 ultramicrotome at -110 ºC. The sections were captured on sucrose-PBS solution drops, and transferred to 10 mm glass coverslips (thickness nr. 1.5; Marienfeld Cat#0111500). Sections on sucrose drops were stored at -20°C.

2.5 BAC probes

BAC probes were obtained from BACPAC Resources (Oakland, CA, USA). BAC Probes were labelled with AlexaFluor 488, 568 and 647 by nick translation according to the manufacturers's instructions (Roche), and separated from unincorporated nucleotides using MicroBioSpin P-30 chromatography columns (BioRad, Hertfordshire, UK).

Table 2. 1 – List of BAC clones used
All BAC clones were obtained from BACPAC Resources (Oakland, CA, USA) and validated by PCR of specific sequences within the corresponding genomic region. Genomic coordinates are in mm9.

Clone	Species	Chromosome	Genomic region	Fluorophores
RP24-312A1	Mouse	11	33330710 - 33520509	Alexa 488
RP23-222N17	Mouse	11	41612930 - 41811642	Alexa 647
RP23-216H23	Mouse	11	52129278 - 52331936	Alexa 488, Alexa 568, Alexa 647
RP23-15E2	Mouse	11	62285129 - 62487781	Alexa 568

2.6 Design of myTag probes

The myTags FISH probes were ordered from the company Arbor Biosciences (Ann Arbour, USA) and designed with help of Kassandra Semrau. MyTag probes consist of libraries of oligos that cover individual target regions of 43-47 length, that are each labelled with a fluorophore. Probe design was adapted to ensure maximum specificity of the probe libraries to the target sequence. Fluorophore density was designed for maximum signal intensity, whilst also making it comparable amongst different probes used in one experiment, to ensure similar signal intensity. The design process of myTag probes is shown in Figure 2.1.

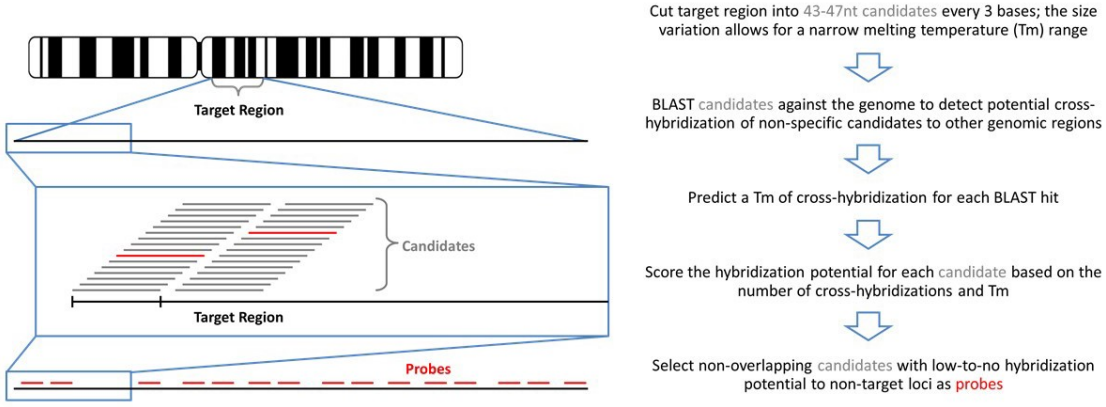


Figure 2.1: Design of myTag probes

MyTag probes are designed to ensure maximum specificity, by eliminating potential cross-hybridization. (image taken from myTags: http://www.arborbiosci.com/products/pre-labeled-libraries/).#

Table 2.2 – List of myTag probes used

Listed below are FISH probes used for 3D and cryoFISH experiments to investigate the genome architecture of the Sox9 Locus, as well as the organization of several super-enhancers. The probes are MyTag probes and were obtained from the company Arbor Biosciences (Ann Arbour, USA). Coordinates are indicated in mouse genome assembly mm9. Some libraries were labelled separately with one or two different tags to allow flexible co-hybridization with pairs or triplets of probes.

Probe #Reference	Species	Chromosome	Genomic region	Tag
SE1 (C3D63S)	Mouse	3	87930000-88430000	ATTO 550; ATTO 594
SE2 (C3D70S)	Mouse	3	96320000-96820000	ATTO 647; ATTO 550
SE3 (C3D75S)	Mouse	3	102580000-103080000	ATTO 594
Low1 (C3D65L)	Mouse	3	91070000-91570000	ATTO 488; ATTO 647N
SE4 (10429)	Mouse	11	33280000-33780000	ATTO 594; ATTO 550
SE5 (10431)	Mouse	11	52150000-52650000	ATTO 550; ATTO 594
SE6 (11241)	Mouse	11	61880000-62380000	ATTO 647N ATTO 488
Low2 (10432)	Mouse	11	41400000-41900000	ATTO 488 Biotin
A	Mouse	2	170294745-170714745	6FAM ATTO 550
Kcnj TAD	Mouse	11	110486662 - 111311965	ATTO 550
Sox9 A	Mouse	11	111511128 - 112514692	ATTO 488

Sox9 B	Mouse	11	112514693 - 112927683	ATTO 594

Table 2. 3 –MyTag top and bottom super-enhancer probes
The two FISH probes, top and bottom interacting super-enhancer FISH probes shown below each consist of 18 labelled regions. Both are MyTag probes and were obtained from the company Arbor Biosciences (Ann Arbour, USA). Coordinates are indicated in mouse genome assembly mm9.

2.7 Fluorescence in situ hybridization

Top interacting super-enhancers FISH probe

mouse, mm9, Atto594

Chromosome	Genomic region
Chr1	138240000-138640000
Chr1	154600000-155000000
Chr1	121000000-121400000
Chr2	152440000-152840000
Chr2	30840000-31240000
Chr2	71400000-71800000
Chr3	102800000-103200000
Chr3	135000000-135400000
Chr3	88640000-89040000
Chr4	123000000-123400000
Chr4	136840000-137240000
Chr4	98200000-98600000
Chr7	56320000-56720000
Chr7	119640000-120040000
Chr10	66280000-66680000
Chr10	20680000-21080000
Chr11	8720000-9120000
Chr11	33040000-33440000

Bottom interacting super-enhancers FISH probe

mouse, mm9, Atto550

Chromosome	Genomic region
Chr3	132360000-132760000
Chr3	35120000-35520000
Chr8	87560000-87960000
Chr8	74720000-75120000
Chr9	120040000-120440000
Chr9	56320000-56720000
Chr9	78120000-78520000
Chr12	12880000-13280000
Chr12	55400000-55800000
Chr15	25520000-25920000
Chr15	60920000-61320000
Chr15	97000000-97400000
Chr16	84600000-85000000
Chr16	22720000-23120000
Chr16	29600000-30000000
Chr19	25320000-25720000
Chr19	5480000-5880000
Chr19	53280000-53680000

2.7.1 Preparation of FISH probes

BAC probes were co-precipitated with mouse Cot1 DNA (Roche; 0.2-0.8 µg/µl final concentration). myTag probes did not require addition of Cot1 as they are specific to non-repetitive regions of the genome. DNA was precipitated in 100% ethanol and 3M sodium acetate and resuspended in hybridisation buffer (50% deionised formamide, 10% dextran sulfate, 2xSSC, 50 mM phosphate buffer pH 7.0). Before hybridisation, DNA probes were denatured at 70ºC for 10 min and re-annealed at 37ºC for 30 min.

2.7.2 Cryo-FISH

DNA FISH was performed as previously described (Ferrai et al., 2010). Briefly, cryosections were first rinsed (3x) in PBS, incubated (15 min) in 20 mM glycine in PBS, rinsed (3x) in PBS, permeabilized (10 min) with 0.2% Triton X-100 in PBS, and then washed (3x) in PBS. Cryosections were incubated (1 h, 37 °C) with 250 µg/ml RNase A (Sigma; in 2xSSC), treated (10 min) with 0.1 M HCl, dehydrated in ethanol (50 to 100% series, 3 min each), denatured (10 min, 80 °C) in 70% deionized formamide, 2xSSC, and dehydrated in ethanol as before. Hybridization was carried out at 37 °C in a moist chamber over 48h. Post-hybridization washes were as follows: 50% formamide in 2xSSC (42 °C; 3x over 25 min), 0.1xSSC (60 °C, 3x over 30 min), and 0.1% Tween-20 in 4xSSC (42 °C, 10 min). Sections were washed in 2xSSC. Nuclei were counterstained with DAPI present in the mounting media (DAPI/VectaShield; Vector Laboratories) immediately before imaging.

2.7.3 3D-FISH

Cells were grown on glass coverslips and fixed (30 min) with 4% PFA in 250 mM HEPES pH 7.8, containing 0.1% Triton X- 100. Fixed cells were permeabilised with 0.5% Triton X-100 containing 20 mM glycine and 0.5% saponin in PBS (10 min). After incubating with 0.1 M HCl (10 min), cells were equilibrated in 20% glycerol in PBS (20 min), and subsequently subjected to 1-4 freeze/thaw cycles using liquid nitrogen. Coverslips were stored at -80ºC after the first cycle. After defrosting, and washing in PBS to remove glycerol, cells were post-fixed in 1% PFA in PBS (10 min) before being heat-denatured (80ºC, 8-10 min) in 70% formamide in 2xSSC/50 mM phosphate buffer (pH 7.0). Cells were rinsed in 50% formamide in 2xSSC before hybridisation with the respective DNA probes for >40 h, at 37ºC. Post-hybridisation washes were as for cryo-FISH. Nuclei were counterstained and coverslips

mounted as before.

2.8 Immunofluorescence stainings of cryosections and 3D cells

For immunolabelling of whole cells, ES cells were fixed as for 3D-FISH. Cryosections or whole cells were washed in PBS, permeabilised with Triton X-100 in PBS (0.5%, 30 min, for whole cells; 0.1%, 10 min, for cryosections), 20 mM glycine in PBS (20 min) and blocked for 1 h with PBS+ (1% BSA, 0.1% casein, 0.2% fish skin gelatin, in PBS, pH 8.0). All washes and antibody dilutions were done with PBS+. Primary antibodies (Table 2.4) were incubated for 2 h, followed by a 1 h wash, and secondary antibodies (Table 2.5) were incubated for 1 h, followed by a 30-45 min wash. For immuno-cryo-FISH, antibodies were fixed with 8% paraformaldehyde in 250 mM HEPES pH 7.6 (1 h) prior to FISH.

The immunostaining method was adapted from Ferrai et al. (2010). The ultrathin cryosections were washed three times in PBS, 0.05% Tween-20 (v/v), each step 10 min, followed by incubation for 15 min in 20 mM glycine (w/v) (Sigma-Aldrich® Cat#68898), 0.05% Tween-20 (v/v) in PBS solution. The cryosections were permeabilized for 10 min in 0.1% Triton X-100 (v/v), 0.05% Tween-20 (v/v) in PBS. The residual detergent was removed by rinsing multiple times in blocking solution (1% BSA (w/v), 0.05% casein (w/v), 0.2 % gelatin from cold water fish skin (v/v), 0.05% Tween-20, in PBS, pH 7.6). The cryosections were incubated for 1 hour in blocking solution at room temperature. Primary antibodies (Table 2.4) were incubated for 2 h, followed by a 1 h wash, and secondary antibodies (Table 2.5) were incubated for 1 h, followed by a 30-45 min wash. For immuno-cryo-FISH, antibodies were fixed with 8% paraformaldehyde in 250 mM HEPES pH 7.6 (1 h) prior to FISH.

Table 2. 4 – List of primary antibodies used

Antibody	Host organism	Company
Oct4	Mouse anti-Oct4, clone 40/Oct -3	BD biosciences
Nanog	Rabbit anti-Nanog, Ab80892, polyclonal	Abcam
Splicing speckles	Human anti-Sm	Jackson
Pan-histones	Anti-Histone, clone H11-4	Milipore

Table 2. 5 – List of secondary antibodies used

Secondary Antibody	Host organism	Company
AlexaFluor 647 *Anti human IgG*	Goat	Invitrogen (Thermo Fisher)
AlexaFluor 488 *Anti mouse IgG*	Donkey	Invitrogen (Thermo Fisher)
FITC *Anti human IgG*	Donkey	Jackson ImmunoResearch
AlexaFluor 488 *Anti rabbit IgG*	Donkey	Invitrogen (Thermo Fisher)

2.9 Confocal Microscopy

Images were acquired on a confocal laser-scanning microscope (Leica TCS SP8; 63x oil objective, NA 1.4), using pinhole equivalent to 1 Airy disk. Images from different channels were collected sequentially to prevent fluorescence bleed-through. Raw images (TIFF files) were merged in ImageJ and contrast stretched without thresholding in Adobe Photoshop. Image acquisition was done randomly based on the DAPI staining, to avoid bias.

2.10 CryoFISH Image analysis

2.10.1 Estimation of average nuclear radius

Cryosections of mESC 46C cells were prepared, stained with DAPI and imaged on a confocal laser-scanning microscope as described above. The radii of nuclear profiles (NPs) in the sections were measured using an automated macro developed by M. R. Branco (Ana Pombo laboratory). The average nuclear radius of mESCs was computed according to a method previously established by Antonio Scialdone (Nicodemi laboratory, Università di Napoli Federico II, Italy) from the radii of NPs using the formula:

$$R_N = (4/\pi) \cdot R_S$$

where R_N is the average nuclear radius and R_S is the average section radius under the approximation of spherical nuclei (Beagrie, PhD dissertation). From the average nuclear Radius, the average nuclear volume is calculated with the formula:

$$V_N = 4/3 \cdot \pi \cdot (R_N)^3$$

Where V_N is the average nuclear volume.

2.10.2 Centre of mass distances of FISH signal

Center of mass distances between cryoFISH signals were determined using an ImageJ based script written by Miguel Branco (Ana Pombo laboratory); modifications to improve nuclear segmentation by manual thresholding were implemented by Tiago Rito (Ana Pombo laboratory). Suggested threshold is the maximum entropy threshold. Selection is confirmed after visual inspection. Distances between the centers of mass are determined by triangulation of 2D distances. The complete macro script can be found in the Appendix, Figure 7.3.

2.10.3 Edge-to-edge distance of FISH signal

Edge-to-edge distances were calculated on a Python based edge detection script written by Robert Beagrie (Ana Pombo laboratory, MDC Berlin). The script calculated closest edge-to-edge distances of previously segmented images (macro described in 2.12.2) between loci of different colour in the same nucleus.

2.10.4 FISH probe volume measurement

Volume analysis of cryoFISH probes was done by first segmenting images and subsequently using ImageJ analyse particles, to obtain the signal area. Figure 2.2 details the volume analysis. To avoid bias, when area comparison was performed within different samples of images, segmentation was done with the maximum entropy threshold and subsequent visual inspection of signal encompassed by threshold. The maximum entropy threshold was also the default segmentation of the center of mass macro utilized for analysing the inter-probe FISH distances (see above, 2.10.2.). Area of selection is given in square pixels, or square units (e.g. nm^2), when image dimensions have been used before to set the scale.

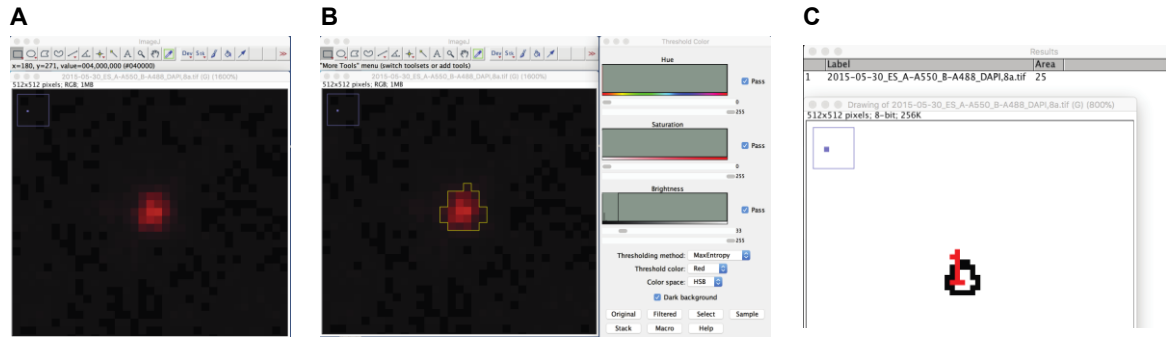


Figure 2.2: Segmentation of particles and subsequent area analysis
A FISH signal in Image as viewed in ImageJ.
B Selection based on the maximum entropy threshold.
C Particle analysis output gives the area in number of pixels and a mask with the labelled outlines of the FISH signal

2.11 3D FISH measurements

2.11.1 3D Image analysis pipeline

3D distance analysis was done using a Matlab script written by Till Hülnhagen (Niendorf laboratory, MDC Berlin). The script is explained in detail in chapter 3.3.2, the complete script can be found in the Appendix section (Figure 7.1). The script analyses the RGB channels individually. To remove the background signal, a 3D median filter was used and subsequently a channel specific threshold was applied, where threshold levels were set based on visual inspection of FISH signal.

FISH signal volumes were obtained by analysing the *regionprops* of the respective signal, which returns the actual number of voxels in the region, which was afterwards multiplied with the voxel dimensions (Figure 3.5C). To obtain the center-of-mass distances between adjacent signals, the center-of-mass of the signal was first determined as x, y and z coordinates (*Centroid*) and then the distance to the nearest neighbour of the other colour calculated. To obtain the border-to-border distances of adjacent FISH probes, a shell was created containing the outer rim of the FISH signal. Using the two shells of for example neighbouring red and blue FISH signals, the distance of each red shell voxel was calculated to each blue shell voxel. The actual border-to-border distance was determined as the minimal distance between any two voxels of the shells.
To determine the overlap between two FISH signals, signals with a border-to-border distance of zero were considered and their overlapping region divided by the area of the smaller region.

2.11.2 3D FISH Volume analysis using ImageJ

An alternative approach to determine 3D volumes of FISH applied to determine 3D cluster volumes of super-enhancer probe multimers was the application of the ImageJ 3D object counter on segmented 3D stacks. Segmentation of stacks was done based on visual inspection. Output voxels were multiplied with the voxel dimensions to obtain the volume. To ensure reproducibility of volumes between using the Matlab pipeline and ImageJ, a comparison between two stacks of images was done and results showed a 3% - 6%, volume difference, where volumes obtained by the Matlab pipeline were of smaller values than values obtained by the ImageJ pipeline (Appendix 7.4).

2.12 Published Datasets

Table 2. 6 - Published datasets used in this work

Dataset	Reference
mESC GRO-seq	Min et al., 2011
mESC H3K27ac	ENCODE project in E14 mESC
mESC super-enhancers and normal enhancers	Whyte et al., 2013
mESC mRNA-seq	Ferrai et al., 2017

3 Genome organization at the TAD level – the Sox9 locus

3.1 Introduction

Chromosome conformation capture based techniques revealed that topologically associating domains (TADs) are one of the basic units of chromatin organization. Based on Hi-C and 5C, they are defined as units of higher self-interaction compared to regions outside of the TAD (Dixon et al., 2012; Nora et al., 2012). Chromosomes are partitioned into TADs, which have been shown to contribute to higher order structures as well as the functional organization of DNA. Previous studies by Nora et al. (2012), showed with 3D FISH that two equidistant regions within a TAD colocalize more frequently than two equidistant regions separated by a TAD boundary (Nora et al., 2012). TAD boundaries have often been thought to limit chromatin interactions, in particular between enhancers and promoters, to within TADs (Andrey et al., 2017). However, a recently published study based on single cell Hi-C and 3D FISH demonstrated that interactions across boundaries between two adjacent TADs occur frequently (Flyamer et al., 2017). Another study showed that TADs are more prominent in specific stages of the cells cycle (Nagano et al., 2017). A better understanding of neighbouring TAD organization and inter-TAD interactions is therefore important to evaluate the importance of TAD boundaries in restricting chromatin contacts. Understanding the single cell variability of TADs can help to ultimately link local TAD organization to single cell variability in gene expression.

Many targets of enhancers lie within their TADs and disruption of TAD boundaries is associated with enhancer hijacking of nearby genes and concomitant ectopic gene expression (Lupianez et al., 2015; Andrey et al., 2017). Enhancer hijacking can also occur after insertion of a genomic duplication into a chromatin environment (Franke et al., 2016). Duplications and deletions can bring enhancers into the proximity of previously segregated promoters, which can trigger formation of new enhancer promoter contacts, some of which possibly drive genomic diseases. Examples where ectopic gene expression has been linked to genomic duplications have recently been reported at the Sox9 and the Mb loci (Franke et al., 2016; Guo et al., 2016; respectively). Since genomic rearrangements that lead to genomic diseases, frequently occur in humans (Mefford et al., 2008; Brunetti-Pierri et al., 2008; Sharp et al.,

2006), it is important to gain insight about how rearrangements reshape the chromatin environment. In the context of genomic rearrangements, duplications that encompass boundary elements are particularly interesting. Depending on whether duplications occur within TADs or in between, they may lead to the formation of a new separate TAD (neoTAD), or a larger TAD that withholds the original and duplicated genomic regions. Recently, Franke et al. (2016) used population-based 4C-seq and capture Hi-C (cHi-C) to study duplications at the Sox9 locus. In one case of a homozygous duplication that contains the boundary between Sox9 and the neighbouring Kcnj TAD, they identified the formation of a neoTAD in mutant limb buds at embryonic day 12.5 (Franke et al., 2016), but it remained unclear how frequent such conformation is across the cell population, and whether other conformations also occur in the presence of the duplication. It remains unknown, whether a homogenous reorganization of the Sox9 genomic region occurs in all cells bearing the duplication and what effect the duplication has on colocalization and spatial distances between TADs.

The Sox9 locus is an important developmental locus containing the gene encoding transcription factor Sox9. Sox9 is crucial for chrondrocyte differentiation and male sex development, where Sox9 induces differentiation of Sertoli cells and later activates many male-specific genes such as Amh (Kashimada et al., 2010). In bone development, Sox9 is expressed during chondrogenesis in all chondroprogenitors and differentiated chrondrocytes, except terminally differentiated hypertropic chondrocytes (Ng et al., 1997; Zhao et al., 1997). Mutations at the Sox9 locus are causal for Campomelic Dysplasia/Autosomal Sex Reversal, a form of male-to-female sex reversal/hermaphroditism which is associated with skeletal malformations and cleft palate (Knowler et al., 2003).

3.2 Research motivation

In this project, my aim was to study the organization of two neighbouring TADs and how the locus structure changes when a duplication spanning the TAD boundary is inserted into the locus. To this end, we designed FISH probes covering the developmentally important Sox9 locus and its neighbouring TAD containing the genes Kcnj2 and Kcnj16, two potassium ion-channels. Using single cell imaging, I investigated intra- and inter-TAD spatial organization in mouse ESC of the Sox9 TAD and the adjacent TAD with the Kcnj2 and Kcnj16 genes. In

particular, I was interested in investigating single cell variability of TAD folding and the degree of intermingling between neighbouring TADs.

To study the effect of the duplication, I used a mouse ES cell line carrying a homozygous 1.56 Mb duplication across the Sox9/ Kcnj TADs (Dup-L), which was previously described in Franke et al. (2016). Using 3D FISH and cryoFISH, I investigated how the genomic duplication changes the folding of the Sox9 and neighbouring TADs in comparison to wildtype mouse ESC. Furthermore, I investigated whether on a single cell level the neoTAD configuration is the most prominent organization, or whether alternative TAD configurations also arise. The Sox9 duplication mouse cell line (Dup-L line) used in this work was derived by Martin Franke (Franke et al., 2016) and a kind gift of Stefan Mundlos. The Matlab image analysis pipeline used in this chapter was written by Till Hülnhagen (Niendorf laboratory, MDC Berlin) and is available in the appendix (Figure 7.1).

3.3 Results

3.3.1 Choosing genomic regions for *fluorescent in situ hybridization* probes around the Sox9 and Kcnj TADs

The Sox9 TAD and its neighbouring Kcnj TAD are located on mouse Chromosome 11. The locus contains two TADs, one TAD encompasses the Sox9 gene (referred to hereafter as Sox9 TAD) and the other TAD contains two genes Kcnj2 and Kcnj16, encoding potassium channels (referred to hereafter as Kcnj TAD). The large gene desert within the Sox9 TAD, downstream of the Sox9 gene, holds many enhancer elements that regulate Sox9 in different tissues (Gordon et al., 2009; Baojin Yao et al., 2015). Figure 3.1A shows cHi-C and 4C-seq data of the locus in E12.5 limb buds. TAD boundaries are shown as red hexagons. TAD boundaries were defined based on TADs published by Dixon et al., (2012). The exact position of the boundary was refined by taking into account the decrease of the 4C signal (Martin Franke, PhD Dissertation; Franke et al., 2016). The equivalent boundaries in ESC 46C are shown in the Hi-C matrix below (Figure 3.1B). It is possible that this approach has underestimated the length of the Kcnj TAD. The TAD boundary between Sox9 and Kcnj is highly conserved across vertebrates (Martin Franke, PhD Dissertation).

The TAD boundaries coincide both in Hi-C and cHi-C data with the characteristic triangles that demarcate high intra-TAD interactions. The 4C-seq signal drops to low values at the TAD boundary, showing a sharp decrease of contacts behind the boundary. The Hi-C, cHi-C and 4C-seq profiles are all depleted in contacts between the TADs. cHi-C and 4C-seq profiles were provided by Martin Franke, Daniel Ibrahim and Stefan Mundlos, Max-Plank Institute for Molecular Genetics, Berlin.

The positions of the Fluorescence in situ hybridization probes (FISH) are indicated below the 4C-seq profiles, with the red genomic probe covering the Kcnj TAD, the blue and green genomic probes covering the adjacent Sox9 TAD (Figure 3.1A). Kcnj probe location was based on the TAD boundaries of the Kcnj TAD with a ~100 kb gap to the left border and a ~200 kb gap between the red and green probe. FISH probe positions within the Sox9 TAD were based on the Sox9 TAD boundaries, the right border of the green probe was chosen such that it coincides with the coordinates of a Sox9 duplication breakpoint (Dup-L) within a mESC cell line, where the duplication occurs at the intersection of the green and blue genomic probe. The blue probe, covering Sox9 is at a distance of ~100 kb to the right hand side border. More specifically, the 1 Mb green probe covers the enhancer containing gene desert within the Sox9 TAD and the blue probe extends over the Sox9 gene. Fluorescent labels of the FISH probes were, red probe: ATTO550, green probe: ATTO488 and blue probe: ATTO594. Table 3.1 contains the genomic coordinates and length of the three FISH probes.

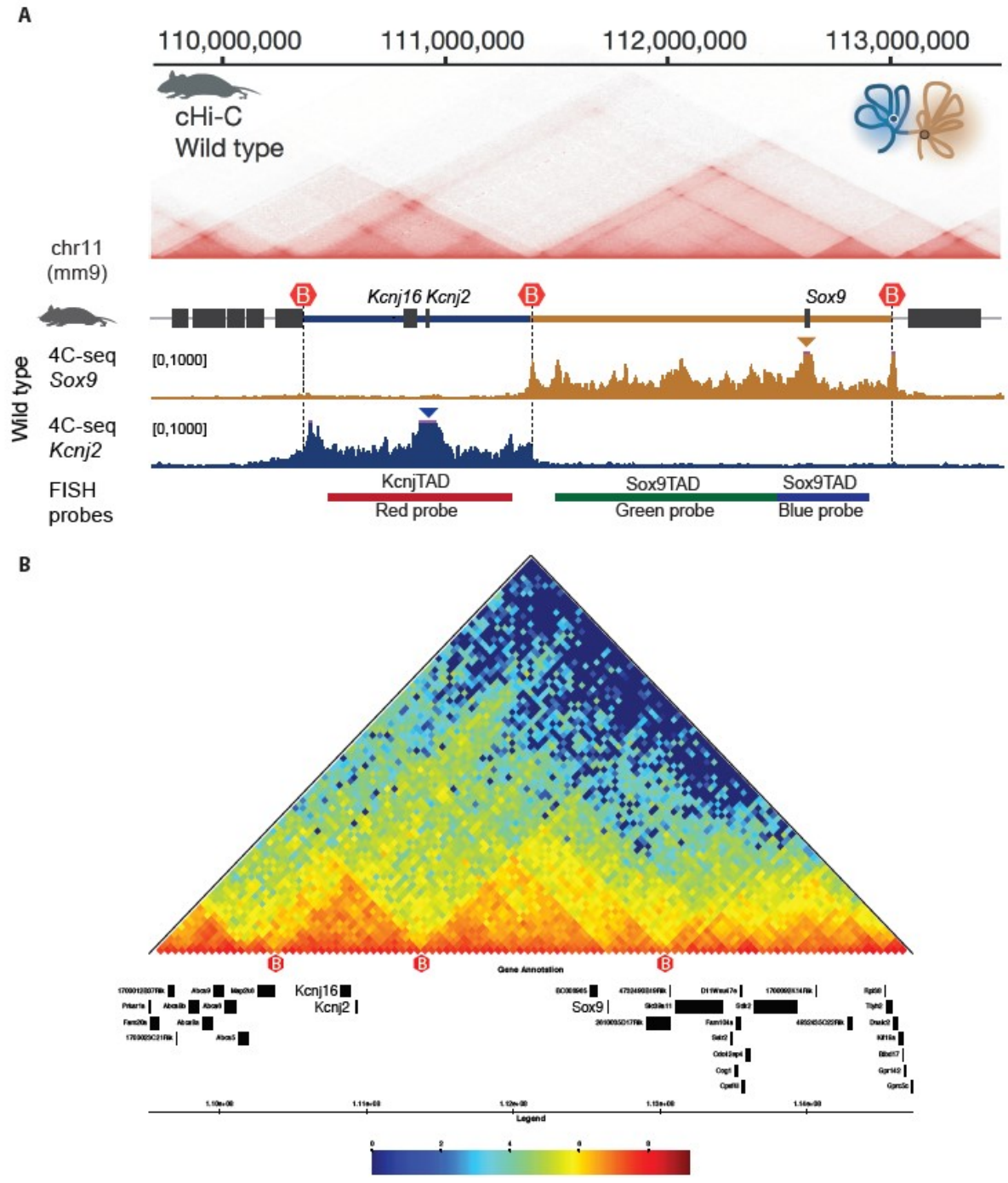


Figure 3.1:The neighboring Kcnj and Sox9 TADs; cHi-C, 4c-seq and Hi-C matrices
A Capture Hi-C (cHi-C) of the Sox9 locus showing separation in two TADs. The TAD boundaries are indicated by the red hexagons. The schematic picture on the top right depicts chromatin folding in Kcnj (blue) and Sox9 (brown) TAD. Below, 4C-seq interaction profiles of viewpoints (triangles) on Sox9 (brown) and Kcnj2 (blue). FISH probe positions are indicated below, the red probe covers the Kcnj TAD, green and blue probes the Sox9 TAD. (Modified from Franke et al 2016)
B Hi-C matrix of the Sox9 locus in mESC. The equivalent boundaries of the cHi-C, 4-seq data are shown by the red hexagons. (Data from Fraser et al., 2016)

Table 3. 1: FISH probes at the Sox9 locus
All FISH probes were obtained from myTag (Arbor Biosciences), probe coordinates are in mm9.

	Fluorophore	Genomic length	Genomic coordinates
Kcnj TAD - red probe	ATTO 550	825,303 bp	chr11:110,486,662-111,311,965
Sox9 TAD - green probe	ATTO 488	1,003,564 bp	chr11:111,511,128-112,514,692
Sox9 TAD - blue probe	ATTO 594	412,990 bp	chr11:112,514,693-112,927,683

The intra-TAD genomic distance between the green and blue probes is 708 kb, as measured from the center of the probes. Across the TAD boundary, distances between the red-green and red-blue probes are 1.11 Mb and 1.82 Mb respectively. Inter-probe distances between the closer intra-TAD probes blue and green are expected to correlate with the high intra-TAD cHi-C interaction frequencies (Figure 3.2). Distances between the further away green and red, blue and red probes, which are separated by the TAD boundary, are expected to be considerably larger with only a small, if any, inter-probe overlap, as interactions across the TAD boundary are expected to be low based on 4C-seq and cHi-C.

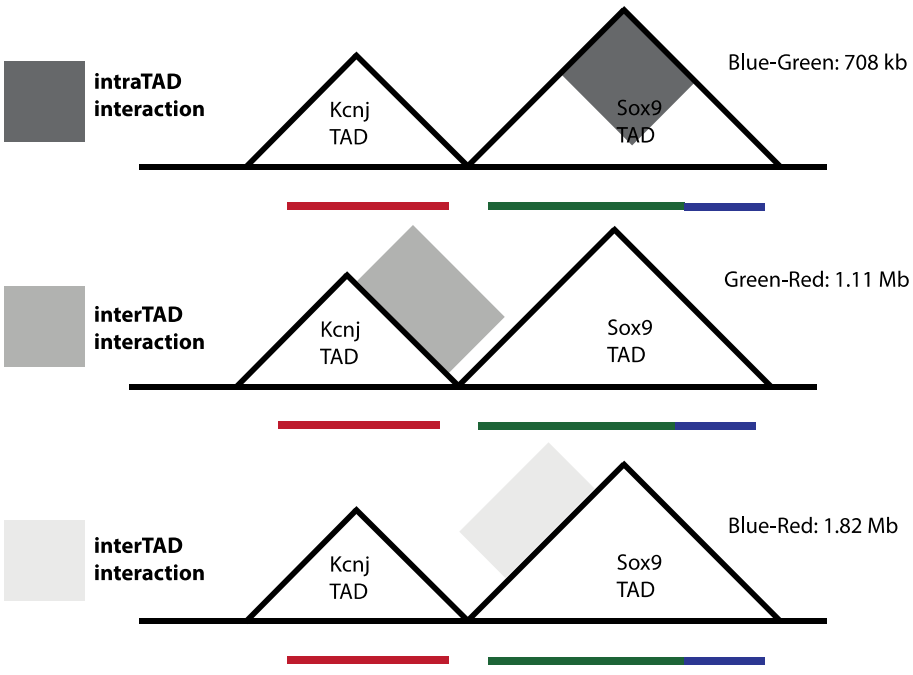


Figure 3.2: Genomic distances between FISH probes
Triangles are schematic representation of the cHi-C triangles, representing the Kcnj and Sox9 TADs. Grey squares indicate the interaction between the respective FISH probes, dark grey symbolizing a strong interaction, light grey a weaker interaction. Linear distances between probes are calculated from center-to-center of the FISH probes.

The expected FISH signal for a strong TAD boundary that prevents inter-TAD contacts, would lead to a scenario where the red signal (Kcnj TAD) is clearly separated from the signal

obtained by the Sox9 TAD (blue and green), as shown in Figure 3.3 on the left. An alternative hypothesis is that TAD boundaries are not strong enough to prevent interactions and instead neighbouring TADs are contacting each other to varying degrees (Figure 3.3, right hand side image).

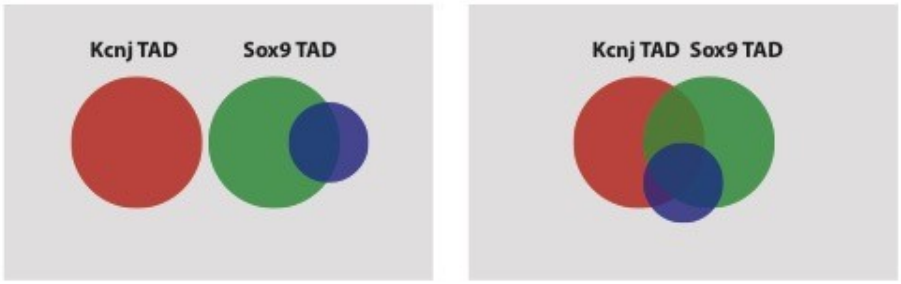


Figure 3.3: Expected FISH signal for a strong TAD boundary compared to a permissive TAD boundary
The picture on the left hand side shows a scenario with a strict TAD boundary. No overlap between the red signal with either blue or green is expected. The strong intra-TAD interactions within the Sox9 TAD are reflected in a high percentage of overlap between the green and blue probes. Alternatively TADs reflect ensemble averages and regions separated by a TAD boundary are often close in space and intermingled.

To visualise the 3D conformation of the Kcnj and Sox9 TADs, I performed 3D-FISH in mouse ESC with the red, green and blue probes simultaneously, and collected 3D stacks of images on the confocal microscope. Figure 3.4 shows two example images of maximum projections of whole stacks. The panels with the magnified signals in Figure 3.4 demonstrate the diversity of 3D locus arrangements that were observed in the cells: (a) are examples of spatial conformations that are in accordance with the proposed model of a clear separation between the two TADs, where the red Kcnj TAD signal is clearly separated from the blue/green Sox9 TAD signal, (b) are examples of conformations where the red signal is located in between the green and the blue intra-TAD signal, indicating a possible stronger inter-, then intra-TAD interaction, and (c) are examples where all three regions localize at equidistance.

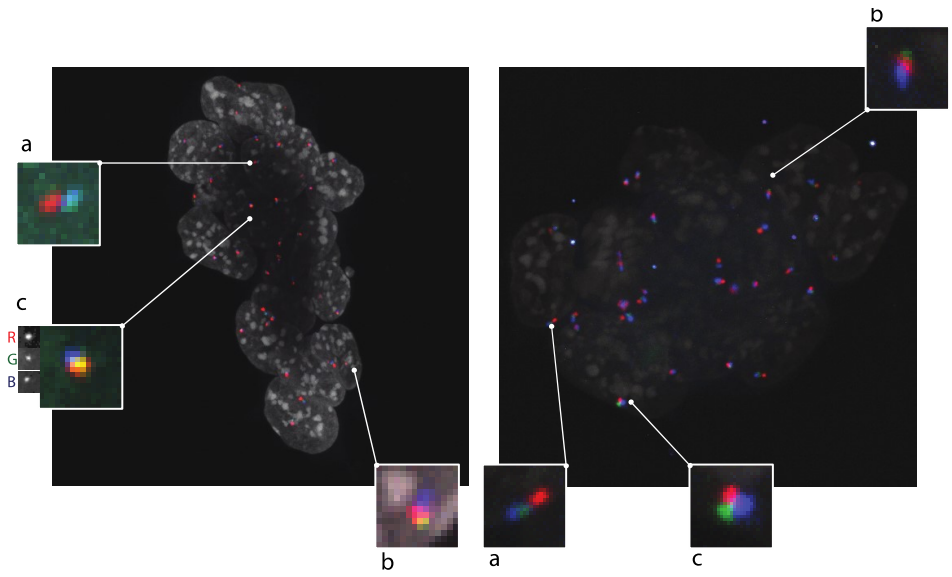


Figure 3.4: Maximum projection of two 3D stacks of ESC cells
The red signal is the Kcnj TAD, blue and green signals the Sox9 TAD, DAPI nuclear counterstain marking the nucleus is shown in grey. The images show clusters of many cells labelled with DAPI. A variety of different neighboring TAD configurations are detected by visual inspection: a) strong intra-TAD inter-action and red neighboring TAD signal is clearly separated (classic TAD image), b) red Kcnj TAD is in between green and blue signal of the Sox9 TAD, c) all three regions are overlapping to the same extent. All 3D ESC imaging data presented here was obtained from 8 3D z-stacks imaged from one hybridization experiment with ESC cells.

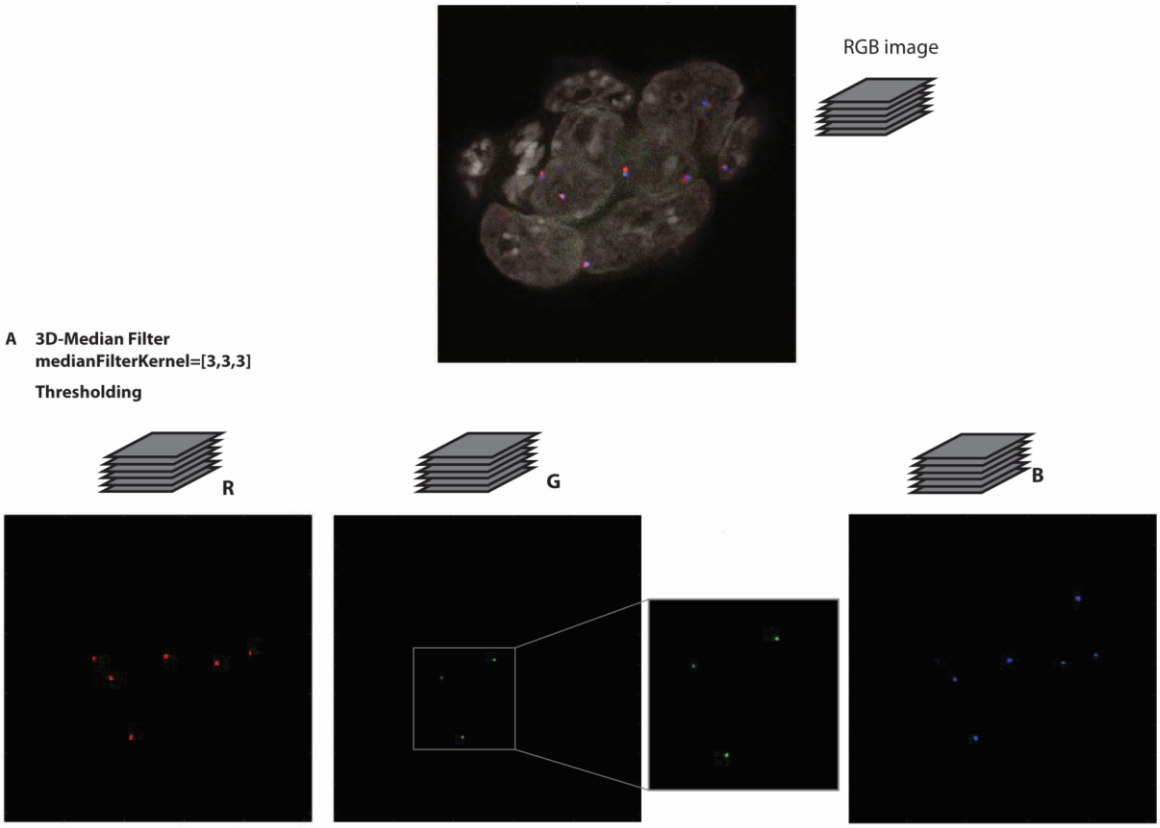
3.3.2 3D image analysis

To obtain a quantitative measure of the 3D imaging data, a MATLAB image analysis pipeline was developed in collaboration with Till Hülnhagen (Niendorf laboratory, MDC Berlin). The complete image analysis pipeline can be found in the Appendix section (Figure 7.1). Figure 3.5 outlines the steps performed by the pipeline. The script analyses the RGB channels individually. To remove the background signal, a 3D median filter was used and subsequently a channel specific threshold was applied, where threshold levels were set based on visual inspection. Examples demonstrating the signal detection after filtering and thresholding compared to the original image are shown in the Appendix (Figure 7.2). The median filter performs median filtering of a 3-D image A in all three dimensions. Each output voxel in the resulting image B contains the median value of the x/y/z neighbourhood around the corresponding voxel in A.

Region property analysis was subsequently performed on the background corrected images. FISH probe volumes were obtained by analysing the *regionprops* of the respective signal, which returns the actual number of voxels in the region, which was afterwards multiplied

with the voxel dimensions (Figure 3.5C). If after visual inspection of the selection and the original image further small background signal was detected, such as individual spots that did not correspond to signal in the original images, a size exclusion was implemented which could remove small background signal below a certain size. To obtain the center-of-mass distances between adjacent signals, first the center-of-mass of the signal was determined as x,y and z coordinates (*Centroid*) and then the distance to the nearest neighbour of the other colour calculated. To obtain the border-to-border distances of adjacent FISH probes a shell was created containing the outer rim of the FISH signal. Using the two shells of for example neighbouring red and blue FISH signals, the distance of each red shell voxel was calculated to each blue shell voxel. The actual border-to-border distance was determined as the minimal distance between any two voxels of the shells.

To determine the overlap between two FISH signals, probes with a border-to-border distance of zero were considered and their overlapping region divided by the area of the smaller region. Additionally, the three channels could be combined to create a 3D reconstruction of the FISH signal (Figure 3.3G).



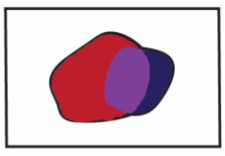
B Probe Volumes

```
% label regions
maskG=(imgGThreshFilt>0);
[lmG, nG] = bwlabeln(squeeze(maskG), 26);
statsG = regionprops(lmG, {'Area','Centroid'});

 % remove regions with area < 10 or area > 1000
for i = 1:numel(statsG)
    if (statsG(i,1).Area<10 || statsG(i,1).Area>1000), lmG(lmG==i)=0; end
end
clear i
% %
% relabel & take voxel dimensions into account
[lmG, nG] = bwlabeln(lmG>0, 26);
statsG = regionprops(lmG, {'Area','Centroid'});
c=struct2cell(statsG);
centroidG=cell2mat(c(2,:));
centroidG=reshape(centroidG,[3,nG]);
areaG=(cell2mat(c(1,:))*voxx*voxy*voxz);
clear c;
```

C Minimum Distance - Center of Mass

```
%% calculate distances between centroids

    scale=[voxx;voxy;voxz];

    nRed=size(centroidR,2);
    nGreen=size(centroidG,2);
    nBlue=size(centroidB,2);

% red to green
    distancesRG=zeros(nRed,nGreen,'single');
    for i = 1:nRed
        for j = 1:nGreen
            distancesRG(i,j)=norm((centroidR(:,i)-centroidG(:,j)).*scale);
        end
    end

%% calculate distances to nearest FISH signal of other
    % red to green
    closestDistancesRG=min(distancesRG,[],2);
    % green to red
    closestDistancesGR=min(distancesRG,[],1);
```

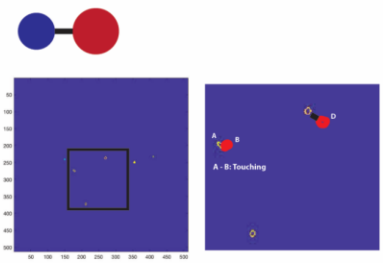
D Overlap



Overlap is defined as the overlapping area divided by the area of the smaller region;

$$\text{Overlap} = \frac{\text{Purple Area}}{\text{Blue Area}}$$

E Minimum Distance - Edge to Edge



F 3D reconstruction

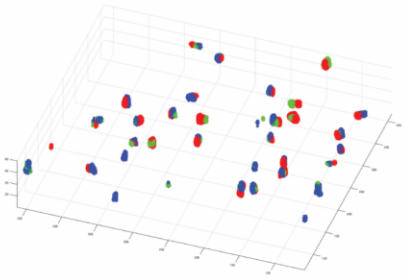
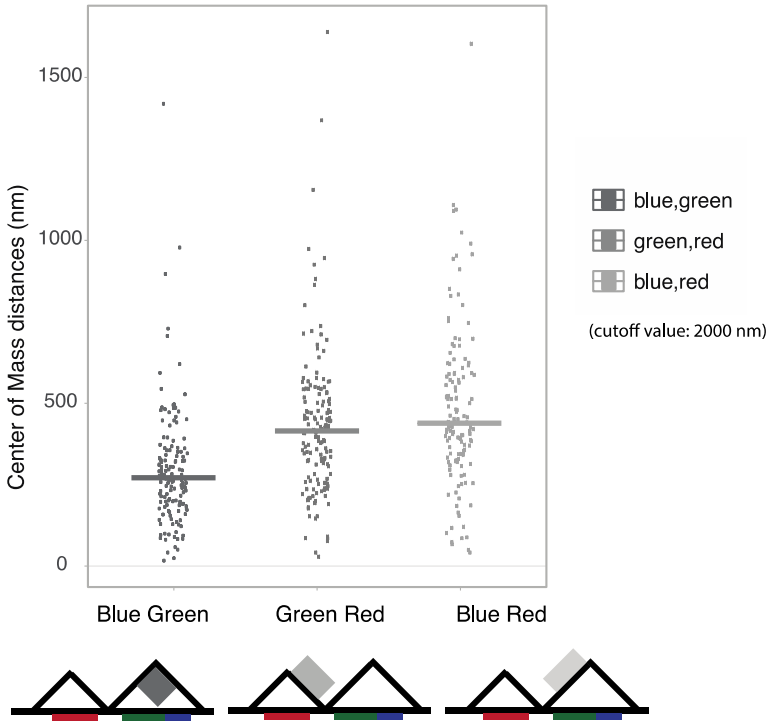


Figure 3.5: Image analysis pipeline to analyze 3D stacks of cells labelled with FISH probes
A,B Thresholding and 3D-median Filter are applied to define the FISH signal.
C Probe volumes – thresholded signal volume is multiplied with voxel dimensions.
D Minimum center-of-mass distances - closest center-of-mass distances between centroids are determined. **E** Minimum border-to-border distance - peripheral signal shell is created and closest pixels to neighbouring signal are the border-to-border distance.
F Percentage overlap of neighbouring FISH probes.
G 3D reconstruction of the FISH signal gives a visual impression of the 3D arrangement of the FISH signal.

3.3.3 The center-of-mass distance between genomic loci within the Sox9 TAD is smaller than across the TAD boundary

To explore the compartmentalization of genomic regions within and between TADs, I measured the physical distances between the red, green and blue genomic regions covering the Kcnj and Sox9 TADs, in 3D mouse ESC. Image analysis was performed using the Matlab pipeline described above (Figure 3.5). The median intra-TAD distance between blue and green regions was 271 nm, with the two probes located 413 kb apart. The median inter-TAD distances were 415 nm (red-green) and 438 nm (red-blue), at inter-probe distances of 1.20 Mb and 1.62 Mb respectively (Figure 3.6). Although the inter-TAD distances are larger than the intra-TAD distances, on many occasions the inter-TAD regions lie within 250 nm (centre-centre-distance), consistent with physical overlap between the TADs. Approximately, 20% of the red-green and 13% of the red-blue distances were below 250 nm, in contrast with 42% of the intra-TAD blue-green distances. This shows that intra-TAD regions are located at a closer median center-of-mass distance, but close spatial proximity between TADs also occurs across the boundary. Therefore, a high degree of single cell variability exists in the compartmentalization of TADs with respect to the locus geometry. This is unexpected, as TADs tend to be regarded as structural units that have high intra-TAD interaction and low frequencies of interaction to regions outside of the TADs.



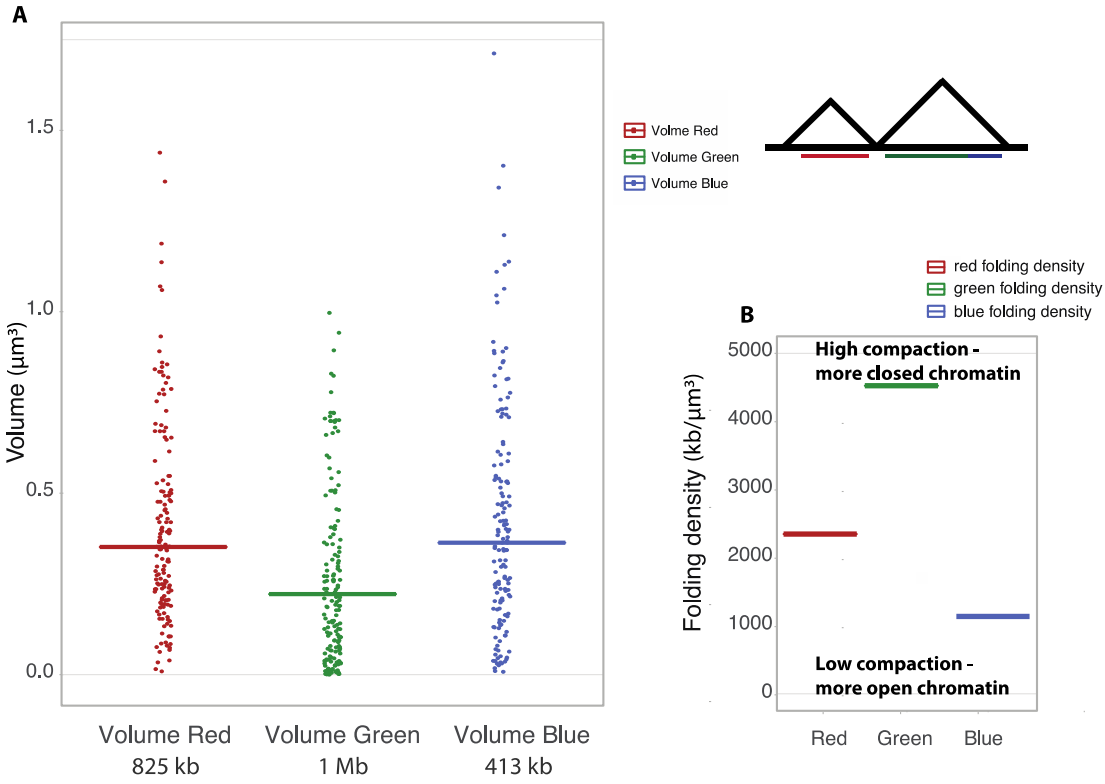
Probes	Number of distances	Median center of mass distance (nm)	Average and standard deviation (nm)	Frequency below 250 nm
Blue Green	133	271	293 ± 893	42 %
Green Red	153	415	428 ± 1553	20 %
Blue Red	134	438	485 ± 1142	13 %

Figure 3.6: Intra-TAD and inter-TAD center-of-mass distances
The plot shows the center-of-mass distances between blue-green, green-red and blue-red Sox9-, Kcnj- TAD FISH probes in mouse ESC. Intra-TAD distances between blue-green are smaller than inter-TAD distances. The table summarizes the number of distances measured, the median inter-probe center-of-mass distances, standard deviation of the center-of-mass distances and the frequency of probes colocalized below 250 nm. A cutoff value of 2000 nm for the center-of-mass distances was applied which is well above the median distance of blue-red with the added value of one standard deviation. Since not all signals were recovered in the image analysis, a cutoff value allows to disregard distances between genomic loci that are not located on the same allele.

3.3.4 The Sox9 and Kcnj TADs show varying degrees of compaction

To further explore the conformation and separation between the Kcnj and Sox9 TADs, I measured their volume. DNA generally exhibits non-uniform folding compaction, which has been linked to the gene expression state of the encompassed region (Boettiger et al., 2016; Zhan et al., 2017). Since the genomic probes cover long stretches of DNA, genomic folding density (kb/μm^3), will significantly influence the volume each region occupies. The probe volume is therefore indicative of the chromatin folding behaviour, which in turn will influence the extent of contact with adjacent genomic regions. In ES cells, the Sox9 and

Kcnj2 genes are Polycomb repressed and the Kcnj16 gene is inactive (Ferrai et al., 2017). The gene desert does not contain genes and is transcriptionally silent in mouse ESC. Figure 3.7A shows the different volumes (μm^3) that Kcnj and Sox9 probes occupy in 3D. Volumes were obtained with the image analysis pipeline (Figure 3.5) by measuring the voxels within the segmented loci. Figure 3.7B shows the folding density (kb/μm^3) of the red, green and blue regions. Interestingly, the volume and corresponding folding density do not directly correlate with the length of the genomic regions labelled. The green probe, despite its length of 1Mb occupies the smallest volume, whereas the smaller 413 kb blue probe is similar in volume to the larger 825 kb red probe. Since the green and blue probes are within the same Sox9 TAD, different folding densities are an indication that different parts of the TAD have different folding properties, possibly allowing them a more autonomous regulation. The Sox9 TAD as a whole is therefore not homogenous in its folding based on 3D FISH. Interestingly, in Figure 3.1 cHi-C and Hi-C show the formation of a subTAD in the Sox9 region, overlapping with the blue probe.



Probes	Number	Median Volume (µm³)	Mean and standard deviation (µm³)	Folding Density (kb/µm³)	Median Radius (µm)
Red	154	0.352	0.411 ± 0.275	2346	0.438
Green	167	0.222	0.255 ± 0.241	4524	0.378
Blue	156	0.364	0.427 ± 0.320	1136	0.443

Figure 3.7: The different FISH probe volumes and the corresponding folding compaction
A Plotted are the volumes of the red, green and blue FISH probes. The blue probe occupies the largest volume, whereas the green probe is the most compacted.
B The degree of compaction of the red, green and blue probes takes into account the genomic length relative to the volume covered by the probe. The gene desert (green) is most highly compacted and the region that covers the Sox9 gene (blue) consists of more open chromatin.

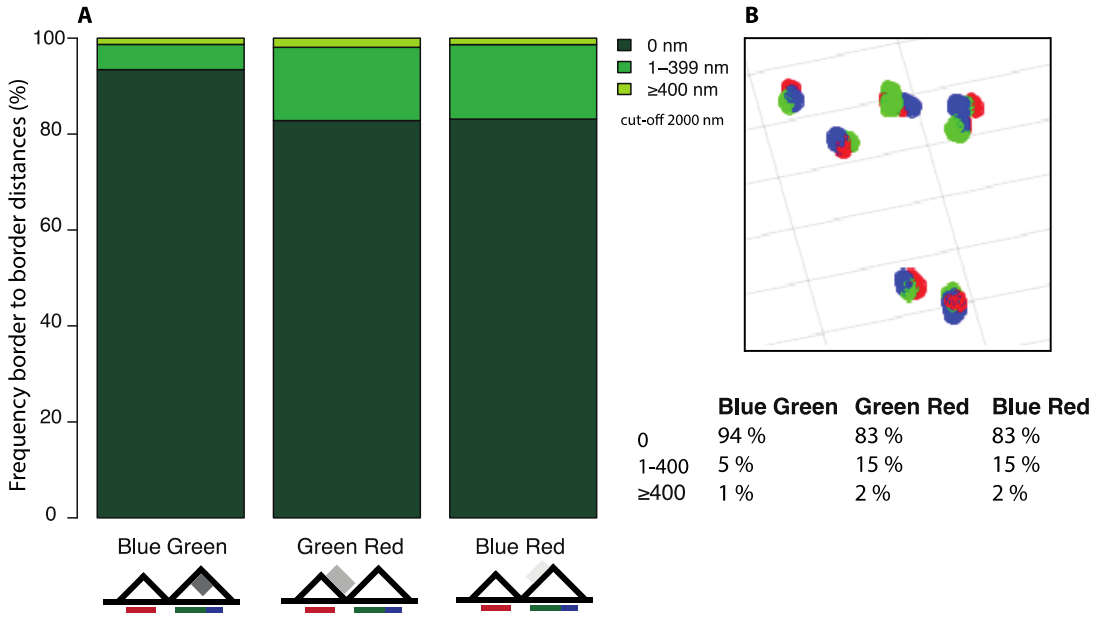
The table contains the number of signals detected, as well as their median volume, standard deviation, folding density and median radius.

To interpret the 3D distances between loci and their spatial relationships, it is important to consider the volumes of the respective probes, as they influence the inter-locus distances between the centres of mass. More voluminous regions may have larger inter-locus center-of-mass distances but can nonetheless contact at their peripheries. I estimated the radii of the red, green and blue regions by assuming sphericity of the 3D volumes marked by the probes. The radii are 438 nm (red), 378 nm (green) and 443 nm (blue). In Figure 3.6, I show that that

intra-TAD distances between green and blue regions within the Sox9 TAD are at a median distance of 271 nm. If the two intra-TAD regions were non-overlapping spheres that are only touching at the periphery, the expected median center-of-mass distance would be the sum of the median radii of the green and blue region, therefore ~820 nm. The discrepancy between expected and measured distance is consistent with significant overlap between the two regions. The same reasoning holds true for the inter-TAD distances, which are detected at 415 nm (red-green) and 438 (red-blue), which is below the sum of their median radii (836 nm and 880 nm respectively). To explore the extent of mixing between intra- and inter-TAD regions, I next considered both the distance between locus borders, and the proportion of overlap.

3.3.5 Border-to-border distances show high spatial proximity of neighbouring regions

The border-to-border distance between two FISH probes is the closest distance between the adjacent signal peripheries. The border-to-border distances between the FISH probes were obtained with the image analysis pipeline described in Figure 3.5. Regions that are overlapping are represented as an inter-probe distance of 0 nm. The spatial proximity between the two TADs is reflected in their close border-to-border distances (Figure 3.8). The high number of intra-TAD probes that are at least touching at the periphery, reflect the high spatial proximity of the regions both within the TAD and across the TAD boundary. 93% of the blue-green adjacent signals are touching or overlapping, 5% are at a border-to-border distance of less than 400 nm and 1% are at a distance above 400 nm. But, 83% of the green-red and 83% inter-TAD blue-red border-to-border distances are at minimum touching at the probe periphery or overlapping. 15% of the inter-TAD signals are at a border-to-border distance between 1 nm and 400 nm. This shows that in fact all three probes are in very close spatial proximity, and the percentage of touching or overlapping regions is only ~11% higher for intra-TAD probes than for inter-TAD probes. To help visualize the range of different configurations, Figure 3.8B shows examples of reconstructed 3D FISH signals, from different cells in the same image (cells are not represented), emphasizing the proximity of the neighbouring regions. Folding is such that proximal TADs are not distinctly separated from each other. Since contacts between neighbouring TADs are very pervasive, it raises the question whether neighbouring TADs only contact at their peripheries or to what extent intermingling of the regions occurs.



	Blue Green	Green Red	Blue Red
0	94 %	83 %	83 %
1-400	5 %	15 %	15 %
≥400	1 %	2 %	2 %

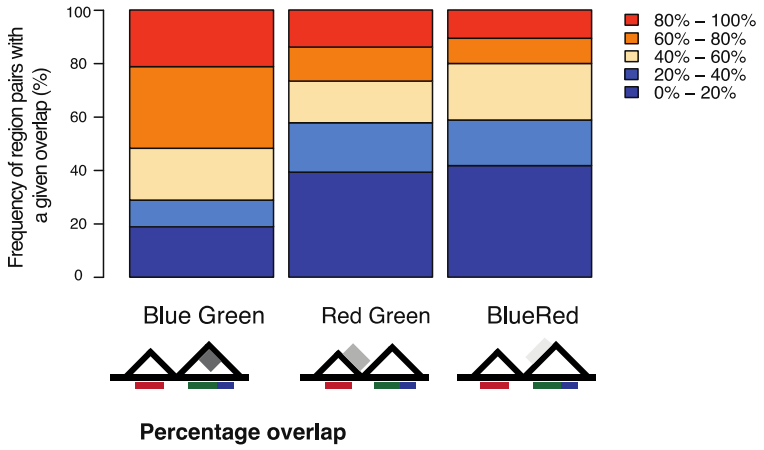
Figure 3. 8: Border-to-border distances between Kcnj and Sox9 FISH probes in ESC
A The stacked bar plot shows the percentage of distances of each blue-green, green-red and blue-red that fall within the categories 0 distance, up to 199 nm above 400 nm.
B Image shows 3D reconstructed FISH signals, which exemplifies the close proximity between the adjacent FISH probes.

The table contains the percentage of inter-probe border-to-border distances falling within the respective distance category.

3.3.6 Extensive intermingling occurs between all regions of the Sox9 and Kcnj locus

To quantify the extent of intermingling between neighbouring regions of the Sox9 and Kcnj TADs, I measured the number of voxels common to any two adjacent red, green or blue regions, and represented them relative to the volume of the smallest region (Figure 3.9); in this way, it is possible to detect when one region is fully enclosed in another, as the percentage of intermingling becomes 100%. These analyses were performed for all regions with a border-to-border distance of 0 nm (in Figure 3.8A). For the intra-TAD regions, more than 70% of the neighbouring green-blue regions showed an intermingling volume of greater than 40% and 21% had an overlap above 80%. Remarkably, regions in neighbouring TADs are also strongly intermingled, with 14% (red-green) and 11% (red-blue) overlapping more than 80% of their volume, and more than half of the neighbouring probes intermingling more than 40% volume. The high degree of intermingling emphasizes the great extent of spatial interaction that occurs on a single-cell level between the neighbouring TADs, despite the intervening boundary. This is in line with recent reports that discovered interactions between

regions separated by a TAD boundary (Flyamer et al., 2017) and suggests that TADs are far from being isolated units, but interact with their neighbouring TADs.



Percentage overlap

	ESC		
	Blue Green	**Red Green**	**Blue Red**
80% – 100%	21 %	14 %	11 %
60% – 80%	31 %	13 %	9 %
40% – 60%	19 %	16 %	2 %
20% – 40%	10 %	19 %	17 %
0% – 20%	19 %	38 %	42 %

Figure 3.9: Percentage overlap between Kcnj and Sox9 TADs compared to Sox9 intra-TAD overlap

The stacked bar plot shows the percentage overlap between neighboring intra-TAD and inter-TAD probes. Probe intermingling is highest between two regions within the same TAD but high overlap also occurs amongst regions that are separated by the TAD boundary.

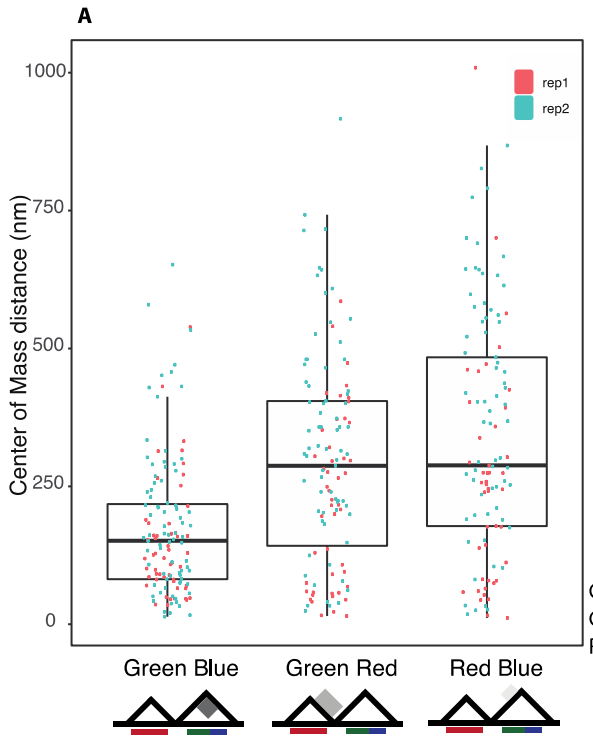
The table contains the percentage overlap of the regions covered by the genomic probes.

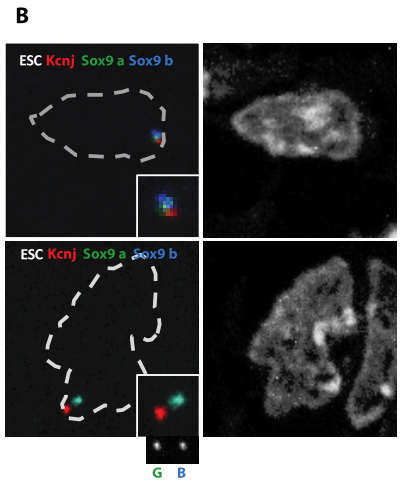
3.3.7 Investigating the distance between the Kcnj and Sox9 TADs with higher resolution cryoFISH

To obtain higher resolution information about the inter- and intra-TAD distances, I performed FISH on mouse ESC cryosections (~200 nm thickness) using the Sox9 and Kcnj TAD probes (Figure 3.10). CryoFISH has the advantages of being performed on cells fixed under conditions that optimize structural preservation, and it improves the axial resolution as all signal detected originates from within the 200 nm thick physical section (Pombo, 2007).

The center-of-mass distances were determined after locus detection and triangulation of 2D distances, using an in house ImageJ macro script that was optimised previously for the use on cryoFISH images, written by Miguel Branco and adapted for manual thresholding of nuclei and FISH signal by Tiago Rito (see Appendix Figure 7.3; Barbieri et al., 2013). Center-of-

mass distances determined using cryoFISH were 158 nm for the intra-TAD green and blue regions, 298 nm and 293 nm for the inter-TAD distances between red-green and red-blue regions (Figure 3.10A). The respective distances determined by 3D-FISH were 271 nm, 415 nm and 438 nm respectively. In parallel colocalization below 250 nm increased, with 77% of the intra-TAD regions in close proximity and 43%, 36% of the inter-TAD regions (green-red, blue-red respectively) colocalized. The difference between the 3D data and the cryoFISH data may be due to partial perturbation of the nuclear structure or expansion of the nuclear volume during the milder fixation and aggressive treatment in 3D-FISH, required to make the nuclear structure accessible to the probes (e.g. permeabilization, freeze-defrost cycles). In cryoFISH, the samples are fixed more strongly fixed compared to 3D cells, making the denaturation step at 80ºC for 3D FISH more susceptible to perturbations. Additionally, cryoFISH imaging gives higher resolution data, due to the better z-resolution. These results show that with high-resolution imaging close proximity and high colocalization between neighbouring TADs can be determined. These results show that neighbouring TADs engage in contacts.





	Median (CoM distances)	Number of distances	Frequency < 250 nm
Green Blue	158 nm	136	77 %
Green Red	298 nm	120	43 %
Red Blue	293 nm	115	36 %

Figure 3.10: cryoFISH data of Sox9-,Kcnj-TAD FISH probes in ESC
A Shown are the pairwise center-of-mass distances between the FISH probes covering the Sox9 TAD and its neighboring Kcnj TAD. Median intra-TAD center-of-mass distances are closer than inter-TAD distances (see table). Two independent FISH experiments were performed for each intra and inter-TAD probe pair (Replicate 1 and 2).
B The images show examples of ESC cryosections with Sox9 TAD (blue and green) and Kcnj TAD (red) probes. Shown separately on the right is the DAPI counterstain. In both images, intra-TAD overlap is stronger than inter-TAD interaction.

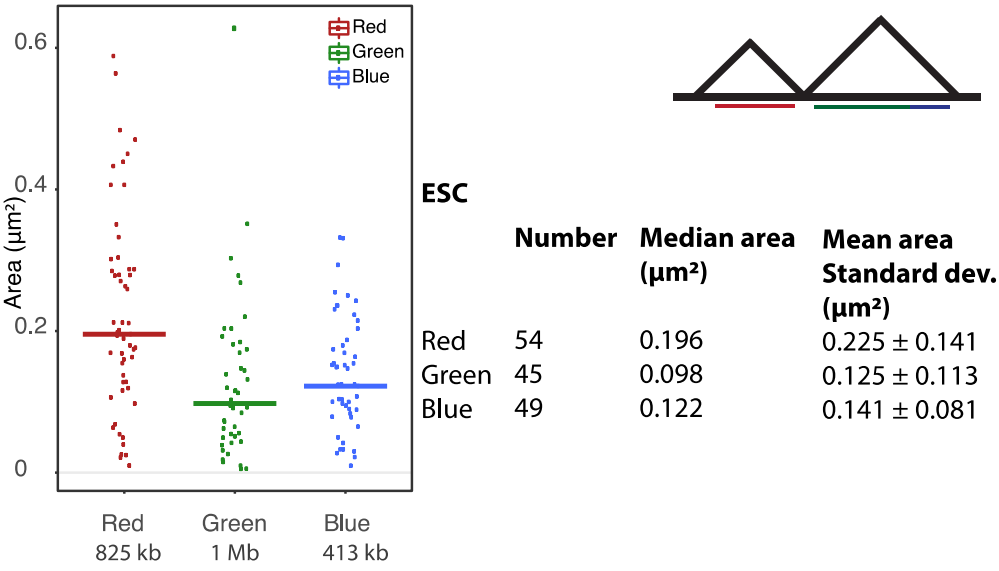
The table contains the median center-of-mass distances, the number of distances measured and the frequency of colocalization below 250 nm.

The observation that the regions at the Sox9 and Kcnj locus are more proximal in cryoFISH raises the question whether also the underlying genomic regions are proportionally more compacted in cryosection nuclei compared to 3D FISH. To determine the areas occupied by cryoFISH signals, segmentation with ImageJ and subsequently the ImageJ particle area analysis function was applied. The area distribution of the three regions is shown in Figure 3.11; the gene desert (green region) remains the region that occupies the smallest area, and therefore is most highly condensed, as seen previously by 3D-FISH.

Both red and blue regions, remain larger than the green region in cryoFISH. The size difference between the red and blue regions is more pronounced, consistent with the fact that

it covers only 413 kb, in contrast to the red probe covering 825 kb. The differences between the 3D volume measurements by 3D FISH and the areas by cryoFISH can be due to differential decompaction level of the red and green probes during the 3D FISH preparation, or alternatively could also be due to a different extent of elongation of the two regions, with a longer ellipsoid region potentially giving rise to smaller areas captured through cryosectioning.

To more directly compare the results from 3D- and cryo-FISH, I estimated the radii of red, green and blue regions by assuming that the regions were circular. The radii measured in cryosections were 250 nm, 177 nm and 197 nm (red, green, blue), in comparison to 438 nm, 378 nm and 443 nm from 3D FISH (red, green, blue). The radii confirm, that all genomic regions occupy smaller volumes by cryoFISH than in 3D FISH. These results raise the possibility of a distortion of the volumes and distances between interacting loci with 3D FISH. The observed differences are of particular interest for the integration of different types of data and methodologies used to study genome architecture.



ESC

	Number	Median area (µm²)	Mean area Standard dev. (µm²)	Median Radius (µm)
Red	54	0.196	0.225 ± 0.141	0.250
Green	45	0.098	0.125 ± 0.113	0.177
Blue	49	0.122	0.141 ± 0.081	0.197

Figure 3.11:Area occupied by cryoFISH probes
The above Figure shows the area distribution as measured in cryoFISH. The red probe occupies the largest area and green the smallest area.

The table contains the number of areas measured, the median area and respective standard deviation, as well as the median radius. Due to time constraints not all images could be analyzed to obtain the probe areas. To obtain cryosection areas, a randomized sample of ~7 images per replicate were analyzed for each red, green and blue channel.

Having studied the structural properties of the Kcnj and Sox9 TADs in ESCs, I was interested in exploring their properties in the Sox9 Dup-L cell line, which bears a duplication of the region that covers part of the Kcnj TAD (excluding the Kcnj genes) and part of the Sox9 TAD (containing the gene desert region that includes Sox9 enhancers, but excludes the Sox9 locus).

3.3.8 The rearrangement of TAD boundaries in congenital disease can lead to 3D alterations and formation of a new TAD

Many duplications have been shown to occur around the Sox9 locus, of which several have severe phenotypic implications in human disease. For example different duplications of a region 0.5 Mb upstream of Sox9 within the so called RevSex domain, a gene desert harbouring regulatory elements, lead to female-male sex reversal (Lybaek et al., 2013). Conversely, duplications that include the RevSex region but extend further upstream result in Cook's disease, a congenital limb malformation characterized by aplasia of nails and short digits. A third type of duplication that also includes the RevSex region, the entire gene desert but not the Kcnj 16/2 genes does not show a phenotype (Franke et al., 2016). All sex reversal associated duplications are located within the Sox9 TAD. Duplications that do not have a phenotype extend into the neighbouring Kcnj TAD (Franke et al., 2016). Using a mouse ES cell line with a homozygous duplication across the Kcnj and Sox9 TAD (Dup-L, Franke et al., 2016), that also includes the TAD boundary, I investigated how this duplication impacts local genome architecture (Duplication: chr11:110,959,589-112,514,692 (1.56Mb)). The cHi-C and 4C-seq profiles of the genomic region including the duplication in embryonic mouse limb buds E12.5 stage, are shown in Figure 3.12 (based on Franke et al., 2016). The black dashed line indicates the duplicated region, which in cHi-C shows to form the pyramidal interaction pattern typical of TADs. The 4C-seq profiles from the duplication breakpoint show interactions restricted to the duplicated region.

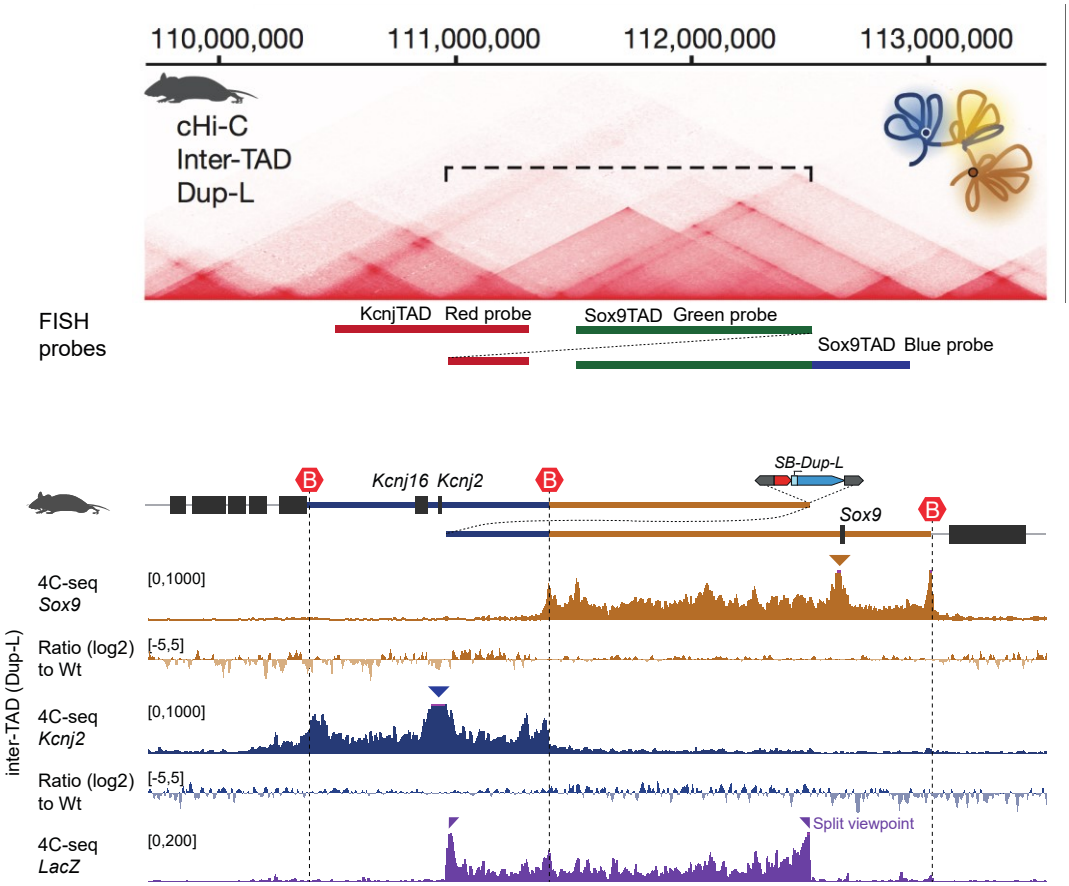


Figure 3.12: The neighbouring Sox9 and Kcnj TADs including a 1.56 Mb duplication covering the boundary

Schematic representation of the Dup-L allele. Position of lacZ reporter at the duplication breakpoint is shown and duplication is indicated by overlap. 4C-seq reads are mapped to the wild-type genome, which results in split lacZ viewpoint. 4C-seq profiles with viewpoint in Sox9 (brown), Kcnj2 (blue) and lacZ reporter (purple) in Dup-L are shown below. Kcnj2 and Sox9 profiles are unchanged, whereas the unique viewpoint in the lacZ reporter shows interactions that are restricted to the duplicated region, suggesting the formation of a separate interaction domain. (modified from Franke et al 2016).

3.3.9 The genomic duplication can lead to different types of 3D arrangements

Before performing experiments, I considered possible 3D reorganisations of the Kcnj-Sox9 locus in the presence of the duplicated Dup-L genomic region in single cells. One possibility is that the duplicated region, which contains a large fraction of the Sox9 TAD, would join the Sox9 TAD forming one large TAD consisting of the original Sox9 TAD together with the duplicated region (Figure 3.13A). The resulting FISH signals that would arise from this fused-TAD scenario would show a single green signal, which should have an increased overlap with the duplicated part of the red Kcnj TAD probe; conversely, the green and blue

regions could show decreased overlap due to the increased length of genomic regions covered by the green probe.

Another possibility, as suggested in Franke et al. (2016), based on cHi-C and 4C-seq data in mouse embryos, is that a new TAD is formed between the duplicated TAD boundary and the original Kcnj-Sox9 TAD (Figure 3.13B). 4C-seq from outside the duplication (Sox9 and Kcnj) shows isolation of the domain, as both domains engage in their wild-type interaction profiles and do not reveal any ectopic contacts. This way, the new TAD boundary would maintain its boundary character and separate the duplicated region from the Sox9 TAD.

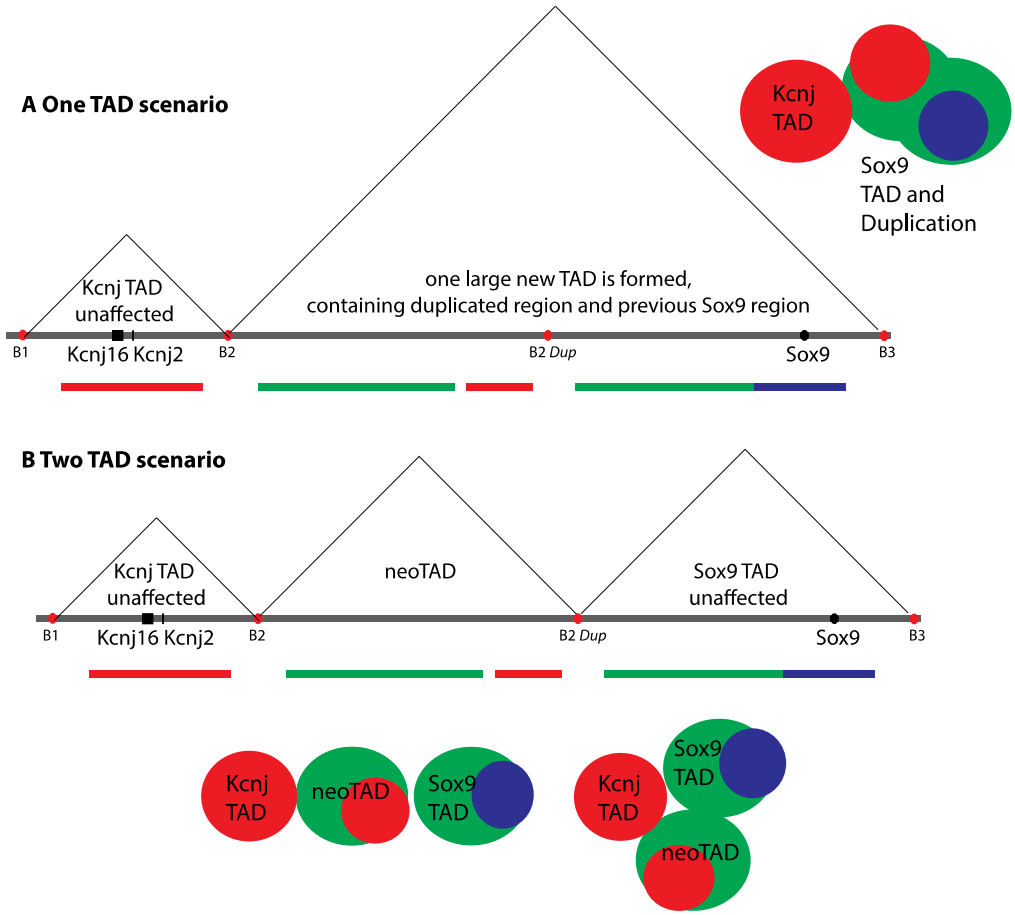


Figure 3.13: Different models for possible rearrangements of the duplicated Sox9 locus
A In the "one TAD scenario" the duplicated region merges with the Sox9 TAD, leading to one large merged TAD.
B In the "two TAD scenario" the duplicated region forms a neoTAD and the Sox9 TAD remains unaffected. Resulting FISH signals are shown by the diagrams.

To test whether a neoTAD structure arises or whether TADs are merged in mESC Dup-L duplication cells and to dissect the single cell organization of the duplicated locus, I performed 3D and cryoFISH on this cell line.

3.3.10 Sox9 Duplication cells maintain stem cell character

To test whether the Sox9 Duplication cells still maintained their stem cell character after cell culture, Oct4 and Nanog immunofluorescence stainings were performed on the 3D duplications cells (Figure 3.14). As in wild type mouse ESC, Oct4 expression is high and Nanog expression shows single cell variability. Nuclei were counterstained with DAPI (pseudo-coloured blue). These immunofluorescence images confirmed the stem-cell character of the Sox9 duplication cells.

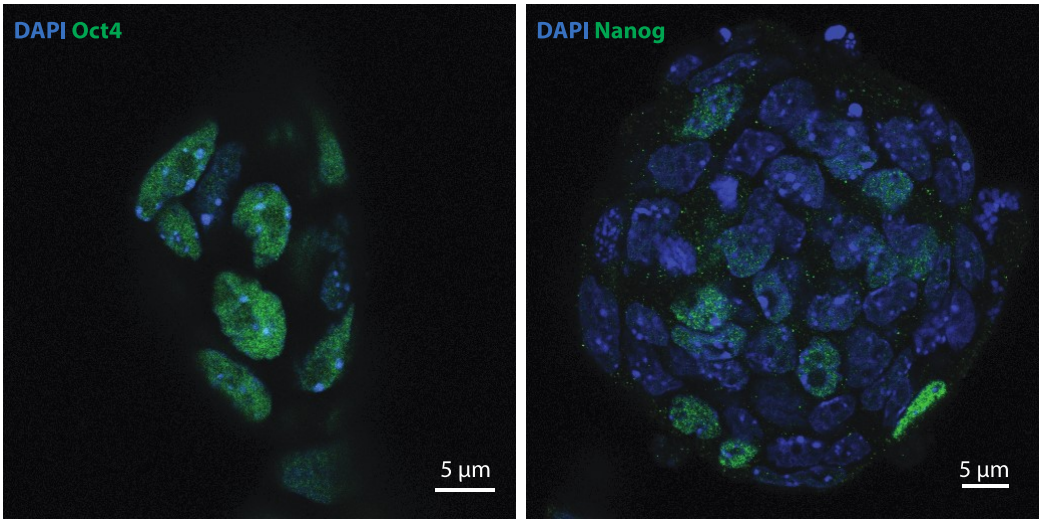


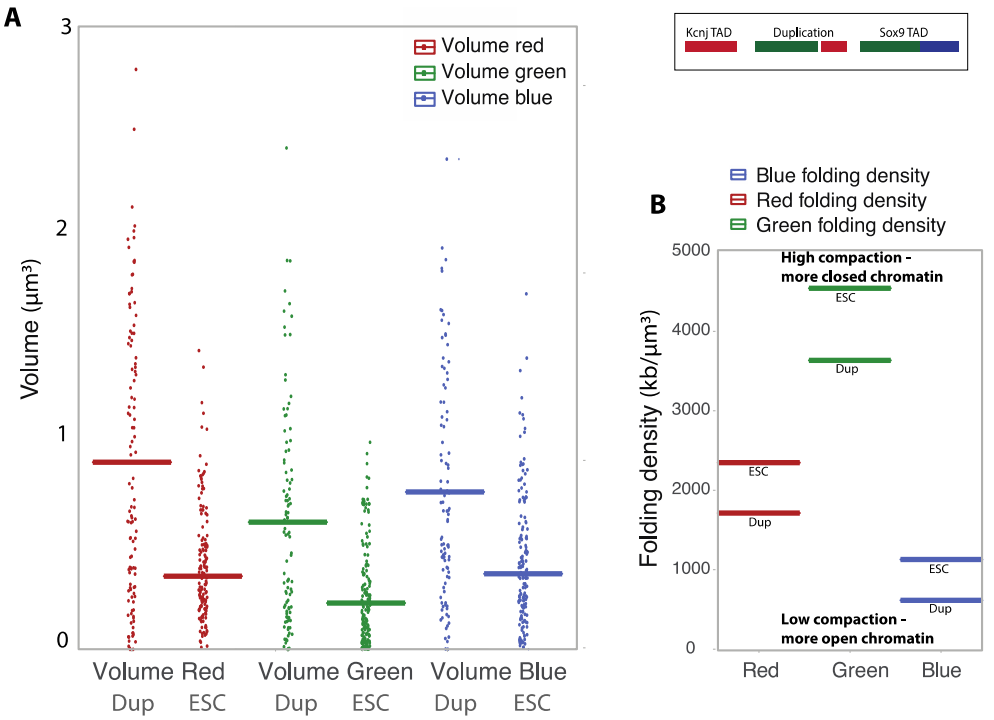
Figure 3.14: Oct4 and Nanog Immunofluorescence staining on Sox9 Duplication cells
The immunofluorescence images show Sox9 Duplication cells, stained with Oct 4 and Nanog antibodies, the DAPI nuclear counterstain is shown in blue. The images confirm expression of the stem cell markers Oct4 and Nanog.

3.3.11 The volume of the Kcnj-Sox9 locus increases for both duplicated and non-duplicated regions

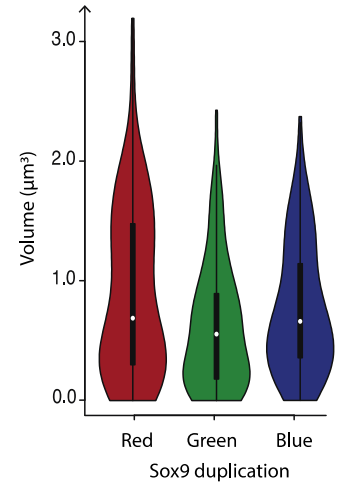
To investigate the conformation of the Kcnj-Sox9 locus in the Dup-L mouse ES cells, I repeated the measurements described in the first part of this chapter, using the same image analyses strategies and criteria. All results presented here were obtained from 15 3D z-stacks imaged from one hybridization experiment with Sox9 Duplication cells, and the same red, green and blue FISH probes. A visual representation of the genomic regions covered by each probe is shown in Figure 3.13. First the number of detected FISH signals and their volumes were measured, to evaluate whether due to the duplication more red and green FISH signals and/or an increase in probe volume could be detected (Figure 3.15). Both the total number of

FISH signals detected in all stacks and the volume of the fluorescent signal, were calculated from the pipeline described previsouly in Figure 3.5.

The total number of duplicated red and green compared to unduplicated blue signals detected in the homozygous Duplication cell line was 114 (red), 102 (green) and 98 (blue), showing that the duplication did not lead to a doubling of the number of red and green signals compared to the blue signals. In the Duplication cell line, only 14% and 4% more of red and green FISH signals were detected compared with blue signals, suggesting that the that formation of neoTAD might be a rarer event than suggested by empirical observation of the cHi-C and 4C-seq data. Next, I measured the volume of the red, green and blue signals. Due to the Dup-L duplication, the genomic coverage of the red and green FISH probes increases from 825 kb to 1177 kb and 1 Mb to 2 Mb, respectively, whereas the coverage of the blue probe remains unaffected by the genomic duplication. Interestingly, I found that the nuclear volumes of the red, green and blue regions were all increased in the duplicated cells compared to ES cells (Fig 3.15). The red region increased from 0.35 μm^3 to 0.69 μm^3 (2.0-fold), the green region from 0.22 μm^3 to 0.55 μm^3 (2.5-fold) and the blue region from 0.36 μm^3 to 0.66μm^3 (1.8-fold). The folding density, which takes into account the increase in genomic length covered by the green and red probes due to the duplication, proportionally decreased (Figure 3.15). Therefore, in 3D FISH the Sox9 duplication locus becomes more decondensed compared to mouse wildtype ESC, suggesting that the duplication not only alters 3D conformation of the locus but also leads to an unexpected significant change in the overall compaction of the duplicated red and green and also the unduplicated blue region. It would be particularly interesting to repeat these experiments in cells that express the Sox9 locus, to explore how the duplication affects the gene expression and 3D folding of the locus in the active state. Together with the increase in volume, the minor change in the detection of red and green signals suggests that the duplicated red region may locate itself in close vicinity to the Kcnj and Sox9 TADs. These observations by 3D-FISH indicate that the formation of a neoTAD may not be the most predominant 3D structure of the Dup-L Kcnj-Sox9 locus in ESC. It is however important to consider that two very closely located but separate domains might not be detected as individual signals by the image analysis pipeline in all cases; which will be later considered in higher-resolution cryoFISH experiments.



C



Sox9 duplication cells (Dup-L)

Probes	Number	Median Volume (µm³)	Mean and Standard deviation (µm³)	Folding Density (kb/µm³)	Median Radius (µm)
Red	114	0.686	0.902 ± 0.701	1716	0.547
Green	102	0.554	0.613 ± 0.520	3624	0.509
Blue	98	0.659	0.759 ± 0.536	627	0.540

ESC

Probes	Number	Median Volume (µm³)	Mean and standard deviation (µm³)	Folding Density (kb/µm³)	Median Radius (µm)
Red	154	0.352	0.411 ± 0.275	2346	0.438
Green	167	0.222	0.255 ± 0.241	4524	0.376
Blue	156	0.364	0.427 ± 0.320	1136	0.443

Figure 3.15: FISH probe Volumes in the Sox9 duplication cells

A Graph represents the volumes (µm³) of the red, green and blue FISH probes. The duplication leads to a volume increase in all three probes, with red showing the largest probe volume and green the smallest. Next to the duplication volumes the median ESC control volume is plotted.

B The folding compaction analysis (volume relative to linear genomic length) in ESC and Sox9 Duplication cells shows how all three genomic regions exhibit a lower folding compaction in the Sox9 duplication cells due to the genomic duplication. The locus therefore becomes more decompacted with the duplication in Dup-L ESC cells.

C Violin plots show the signal volumes in Sox9 Duplication cells. Red probe volumes show a bimodal distribution which might reflect the large Kcnj TAD and the smaller duplicated regions.

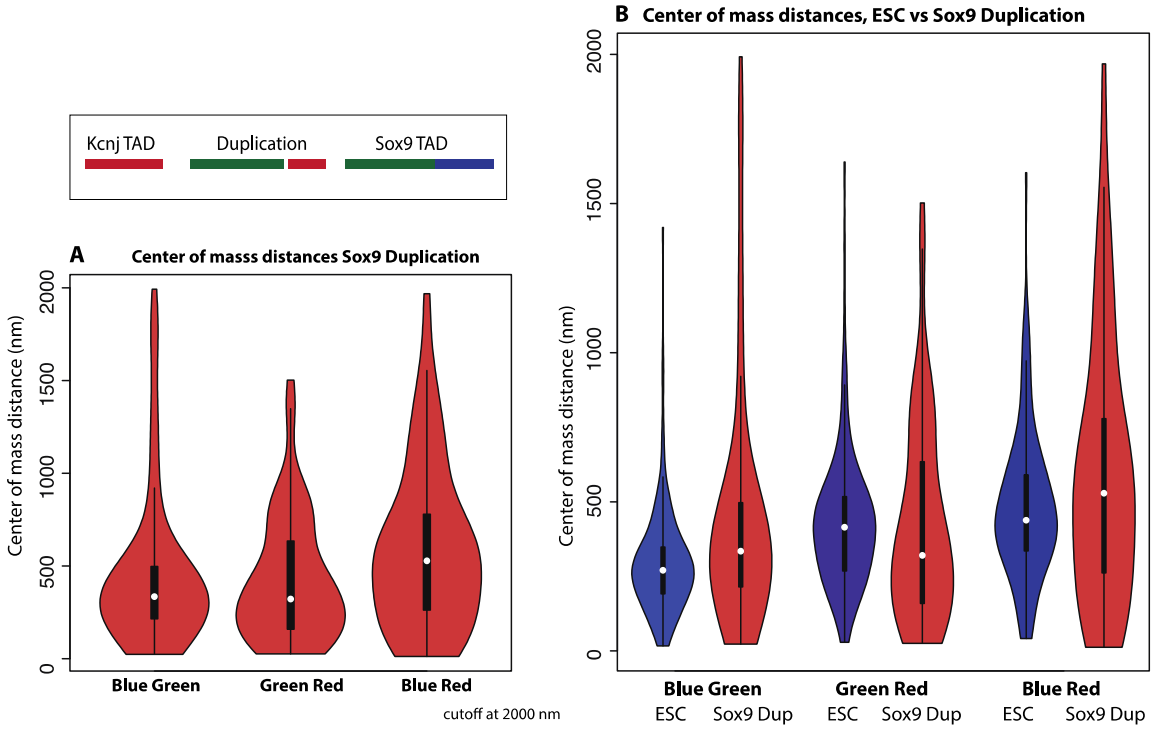
The table contains for both Sox9 Duplication cells and control ESC cells the number of FISH signals detected for the red, green and blue probes, the median volume and its standard deviation, the folding compaction and median FISH signal radius (assuming sphericity).

3.3.12 The center-of-mass distances in the Sox9 Duplication cells are altered compared to the wild type ESC cells

The center-of-mass distances between the red, green and blue probes are a quantitative measure of the locus geometry. Using the image analysis pipeline discussed in Figure 3.5, we can capture the distances between the most closely spaced independent objects labelled in any two colours. Figure 3.16A shows the center-of-mass distances in the Sox9 Duplication cells and 3.16 B the center-of-mass distances in ESC compared to the Sox9 Duplication cells. Distances for each colour are based on the closest distance between two FISH signals of a respective colour. An overall locus reorganization is taking place as can be seen by the changes in inter-probe distances. The blue-green median center of mass distance, which in reflects the intra-TAD distance, increased by 19% to 334 nm. In the duplication, this distance increased, because either in the "one TAD scenario", a large green TAD is formed with a center-of-mass further away from the blue signal. Alternatively, if the duplicated region forms a neoTAD, then in addition to the blue-green intra-TAD interaction an inter-TAD contact between the blue and green area in the neoTAD will be detected, which would lead to an increase in median distance between blue-green. The green-red distance decreased in the Sox9 duplication cells compared to ES cells by 23%, indicating that despite the increased volume of the genomic regions the center-of-mass of the probe regions become more proximal. In the neoTAD scenario, the red-green center-of-mass distance comprises an intra-TAD interaction within the neoTAD and two inter-TAD interactions between the red region in the Kcnj-TAD and the red region in the duplication with the green region in the Sox9 TAD. In the case of one large TAD, the green-red distances contain an intra-TAD distance between the red and green region, and one inter-TAD distance between the Kcnj TAD (red) and the large green region. The distribution of green-red distances in the violin plots appears to consist of two populations, a larger population of FISH probes at a distance below ~700 nm and a smaller population at a distance greater than 700 nm, which could reflect an intra and inter-TAD contact. The blue-red median distance increases by 20% compared to ES cells. The blue-red distance contains in the neoTAD scenario two inter-TAD distances, or in the large green TAD scenario one intra-TAD distance and one inter-TAD distance to the Kcnj TAD.

Interestingly, the percentage of colocalization (<250 nm) for both green-red and blue-red increased in comparison to ES cells, despite the increased center-of-mass distance for blue-

red in the duplication cells. The green-blue centers of mass are less often colocalized. This points to a scenario where 25% of the blue-red distances are in very close proximity, whereas a second population is more distal which would result in an overall increase in median center-of-mass distance, as has been observed. To understand the distance changes in relation to the volume, it is helpful to also consider the border-to-border distances. Genomic probes that are within the same TAD are expected to have a border-to-border distance of zero, whereas inter-TAD interactions are more likely to be at a greater inter-probe distance.



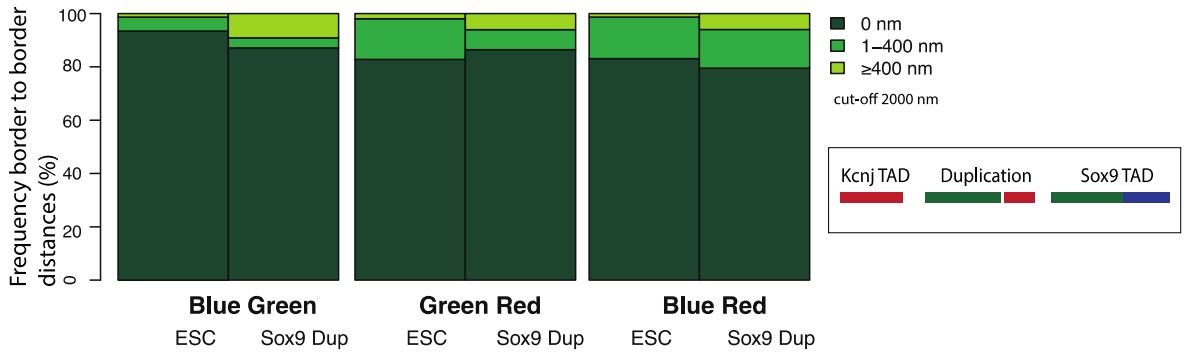
	Probes	Number of distances	Median Center of mass distance (nm)	Mean and standard deviation (nm)	Below 250 nm
Sox9 Duplication	Blue Green	95	334	445 ± 412	33 %
	Green Red	92	321	422 ± 333	37 %
	Blue Red	92	529	593 ± 442	25 %
ESC	Blue Green	147	271	293 ± 893	42 %
	Green Red	153	415	428 ± 1553	20 %
	Blue Red	134	438	485 ± 1142	13 %

Figure 3.16: Center-of-mass distances in ESC and Sox9 Duplication cells
A Shows the center-of-mass distances of the FISH probes in Sox9 duplication cells as violin plots.
B Shows the center-of-mass distances of the FISH probes in ES cells (blue) compared to Sox9 duplication cells (red). The median value is shown as a white dot. Center-of-mass distances increase from ESC to Sox9 duplication cells for both blue green and blue red, for green red the median decreases.

The table contains the number of distances that were analysed for each probe pair, the median center of mass distances of the FISH probes and their standard deviation, as well as the percentage of probes that colocalized below 250 nm.

3.3.13 Changes in border-to-border distances reflect the changed center-of-mass distances

The changes in distances occurring in the Sox9 duplication cells, observed in the center-of-mass distances are reflected in the border-to-border distances. Figure 3.17 shows the distance changes between ES cells and Sox9 duplication cells. 87% of the blue-green border-to-border distances are zero and 13% are at a non-zero distance. This could be a representation of a majority of intra-TAD interactions and a percentage of 13% inter-TAD interactions. Equally, the blue-red border distances are shifted towards slightly less border-to-border overlap with a total of 20% at a non-zero distance. Only the green-red zero border distances increased by ~4%. Since the center-of-mass distance of green-red distances decreased by 23%; the probes within the touching/overlapping category ought to move into closer spatial proximity, which can be evaluated by calculating the percentage intermingling of the FISH probes. It can be summarized, that also in duplication cells all FISH regions covering the Sox9 TAD, the Kcnj TAD and the duplicated region are in very close spatial proximity with a majority of probes that are at minimum contacting at their peripheries. Due to the large volumes of the FISH probes, the intermingling can give a more differentiated picture of the locus geometry, as will be discussed in the next chapter.

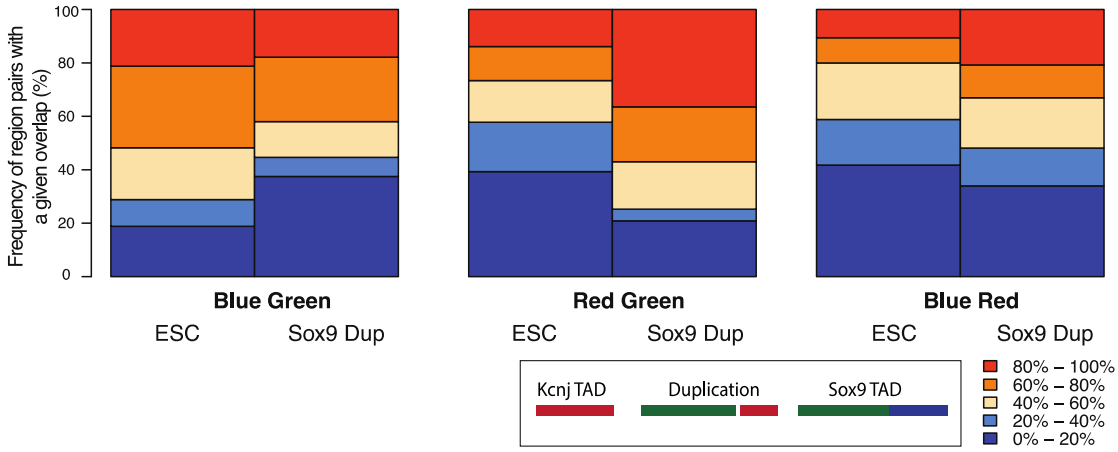


Border to Border distance	ESC			Sox9 Duplication		
	Blue Green	Green Red	Blue Red	Blue Green	Green Red	Blue Red
0 nm	93 %	83 %	83 %	87 %	87 %	80 %
1-400 nm	5 %	15 %	16 %	4 %	7 %	14 %
≥ 400 nm	2 %	2 %	1 %	9 %	6 %	6 %

Figure 3.17: Comparison of the border-to-border distances between ESC and Sox9 Duplication
Shown are the percentage of border-to-border distances, classified as overlapping/touching (0 nm distance), up to 400 nm apart, or at a distance greater then 400 nm. The stacked boxplots compare the border-to-border distances between green blue, green red and blue red in ESC and Sox9 Duplication cells. Distances change between ESC and Sox9 duplication cells showing a slight decrease in overlapping border-to-border distances between blue green and blue red and a slight increase in green red zero nm border-to-border distances.

3.3.14 Changes in the percentage of overlap in 3D Sox9 duplication cells reflect the altered chromatin architecture

Figure 3.18 shows the stacked bar plots for the percentage of the overlap of red-green, green-blue and blue-red. As observed before, the blue-green probes are at a greater distance, which is reflected in the larger number of probes that are overlapping by less than 20%. The green-red probes come into closer proximity, which is visible in the increased percentage of overlap. Amongst the red-green probes in the Sox9 Duplication cells, ~75% are overlapping by more than 40%. Particularly striking is the increase in total overlap of the red-green probes, where 36% are overlapping at a minimum of 80% in Sox9 duplication cells, as opposed to 14% in ESC. The blue-red distances also show an increase in complete overlap, with 21% overlapping at a minimum of 80%. This is in accordance with 25% of the center-of-mass distances being at a distance below 250 nm (Figure 3.16). It is however interesting that the median center-of-mass distance increased. This is likely due to the overall volume expansion, as this most directly influences the center-of-mass distances.



Percentage overlap

NaN corrected: for all NaN = 0

	ESC			Sox9 Duplication		
	Blue Green	**Red Green**	**Blue Red**	**Blue Green**	**Red Green**	**Blue Red**
80% – 100%	21 %	14 %	11 %	18 %	36 %	21 %
60% – 80%	31 %	13 %	9 %	24 %	21 %	12 %
40% – 60%	19 %	16 %	21 %	13 %	18 %	19 %
20% – 40%	10 %	19 %	17 %	7 %	5 %	14 %
0% – 20%	19 %	38 %	42 %	38 %	20 %	34 %

Figure 3.18: Percentage overlap between neighbouring Kcnj-,Sox9-TAD FISH probes in mouse ESC and Sox9 Duplication

The stacked barplots shows the percentage of overlap between the neighbouring intra-TAD and inter-TAD probes in ESC and Sox9 duplication cells. The overlap in blue and green probes is reduced and the overlap in red green and blue red is increased, in particular the percentage of probes that overlap almost fully (80% -100%). The table contains the percentage overlap of the blue green, red green and blue red probes in ESC and Sox9 Duplication cells for five intervals between no overlap (0%) and complete overlap (100%).

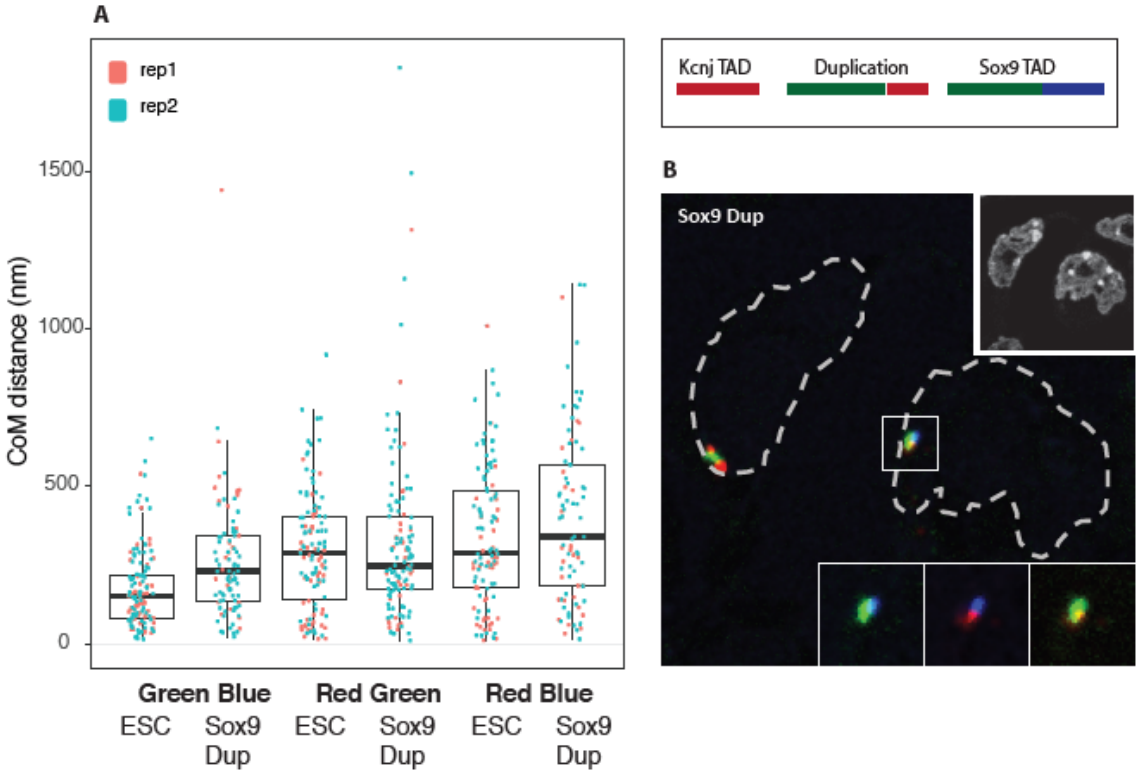
3.3.15 Center-of-mass measured by CryoFISH confirm the 3D results

To obtain high-resolution data I also performed FISH on cryosections of Sox9 Duplication cells. Results are obtained from two independent replicates. All pairwise center-of-mass distances were obtained using an ImageJ macro script (see Appendix, Figure 7.2). To measure FISH probe areas, a randomized selection of images from both replicates were chosen and analysed using ImageJ segmentation and subsequent particle analysis, rendering the FISH signal areas.

Figure 3.19 shows the pairwise center-of-mass distances between the red, green and blue labelled regions, distances are plotted in comparison to wildtype mouse ESC. The green-blue previous intra-TAD distance increased by 34% from 152 nm in ESC to 230 nm in duplication cells. The red-blue median distance increased by 14% in Sox9 Duplication cells to 340 nm. The median red-green distance decreased by 16% from 296 nm in ESC, to 248 nm in duplication cells. CryoFISH shows the same changes as observed by 3D FISH, the local distance changes are therefore reproducible, and the locus rearrangement is consistent using different FISH approaches. Regarding absolute distances in 3D FISH and cryoFISH, the cryoFISH distances were consistently smaller than respective 3D FISH distances.

Figure 3.19B shows an example image of cryoFISH in the Sox9 Duplication cells. The image shows two nuclei, the nucleus on the left is an example of a “one TAD scenario”. The large green signal is part of the new large TAD, and the two red signals on both sides are the Kcnj TAD and the red duplicated region. The nucleus on the right is also a “one TAD” example, the enlarged panels below show the exact signal distribution: The large green signal overlaps with the blue and red signal. Red and blue however do not overlap.

High resolution cryoFISH confirms that the duplication leads to large scale changes in locus arrangement. To evaluate whether also in cryoFISH a decondensation of the locus can be detected, the volume of the areas was measured.



	ESC WT		ESC Sox9 Duplication	
	Median	Number of Distances	Median	Number of Distances
green blue	152 nm	136	230 nm	93
green red	296 nm	121	248 nm	121
red blue	293 nm	115	340 nm	85

Figure 3.19: Comparison of the center-of-mass distances in cryoFISH with FISH probes
A The plot shows the center-of-mass distances for the two replicates in ESC and Sox9 duplication cells between green blue, green red and red blue.
B The image shows an example of a cryoFISH image in Sox9 duplication cells. The circumference of the DAPI nuclear staining is marked by the dashed line, the original DAPI image is shown in the upper right corner. The pairwise green blue, red blue and red green FISH signals of the right nucleus are shown enlarged in the panels below. In this example a large green signal fully overlaps with the red and blue signals, which themselves are touching but not overlapping. The left nucleus contains two separate red signals and a green probe in between.

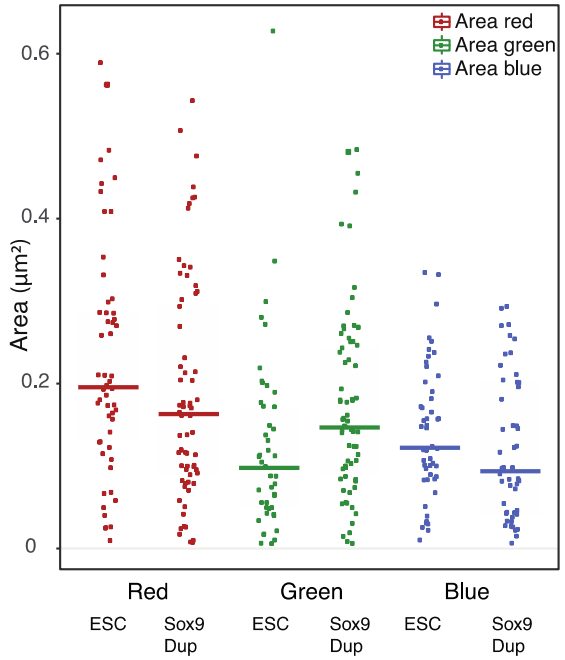
Calculating the probe areas and signal number contributes additional information about changes that occur at the locus and help the interpretation whether a one TAD or two TAD scenario occurs. In 3D FISH, all probe regions showed a volume increase in the Sox9 duplication cells compared to ES cells. In cryoFISH however a different picture emerges. The

median area decreased in red and blue probes by 17% and 23%, respectively from ESC to Sox9 Duplication cell. The green probe showed an area increase by 35% (Figure 3.20). Therefore, whilst 3D FISH suggested that the whole locus expands in volume, cryoFISH results show an increase in compaction for both the red and the blue probe region and an increase in area for the green probe region. The decreased red and blue probe areas perhaps compensate for the increased green probe area, which could keep the total locus volume constant. Differences between the two methods will elaborated in the discussion.

Furthermore, in cryoFISH 27% more red probes and 22% more green probes were detected compared to the blue probe, Figure 3.17C. Also, in mouse ESC cryoFISH, the total number of blue probes was lower than red or green probes, but in the Sox9 duplication cells the detection difference was much higher. The increased number of green probes detected would be consistent with a population of about 22% neoTAD conformations. Volumes displayed in figure 3.17B were calculated from a selection of randomly chosen images from both replicates. The total number of detected signals takes into account all the signals from all the images.

In cryoFISH a differential volume change is seen across all three probes. The green area increased by 35%, whilst its genomic content doubled from 1 Mb to 2 Mb in the duplication. The increased green area favours the interpretation of the formation of a one TAD organization, as two merged green areas would likely lead to an increase of the green area. No doubling of the number of red signals occurs, it therefore suggests, that in duplication cells the locus reorganizes in a way to bring the red Kcnj TAD and the duplicated red region into proximity. The duplication does not affect the blue probe, but the area measured in cryoFISH becomes slightly more compacted in the sox9 duplication cells, suggesting that the underlying genomic region becomes more tightly folded, potentially to accommodate the duplication.

A



B

ESC	Number	Median area (µm²)	Mean area Standard dev. (µm²)	Median Radius (µm)
Red	54	0.196	0.225 ± 0.141	0.250
Green	45	0.098	0.125 ± 0.113	0.177
Blue	49	0.122	0.141 ± 0.081	0.197

Duplication	Number	Median area (µm²)	Mean area Standard dev. (µm²)	Median Radius (µm)
Red	68	0.163	0.188 ± 0.136	0.228
Green	69	0.150	0.171 ± 0.109	0.147
Blue	50	0.094	0.120 ± 0.086	0.094

C **Total number of detected FISH signals**

ESC		Sox9 Duplication	
Red	189	Red	183
Green	191	Green	170
Blue	173	Blue	133

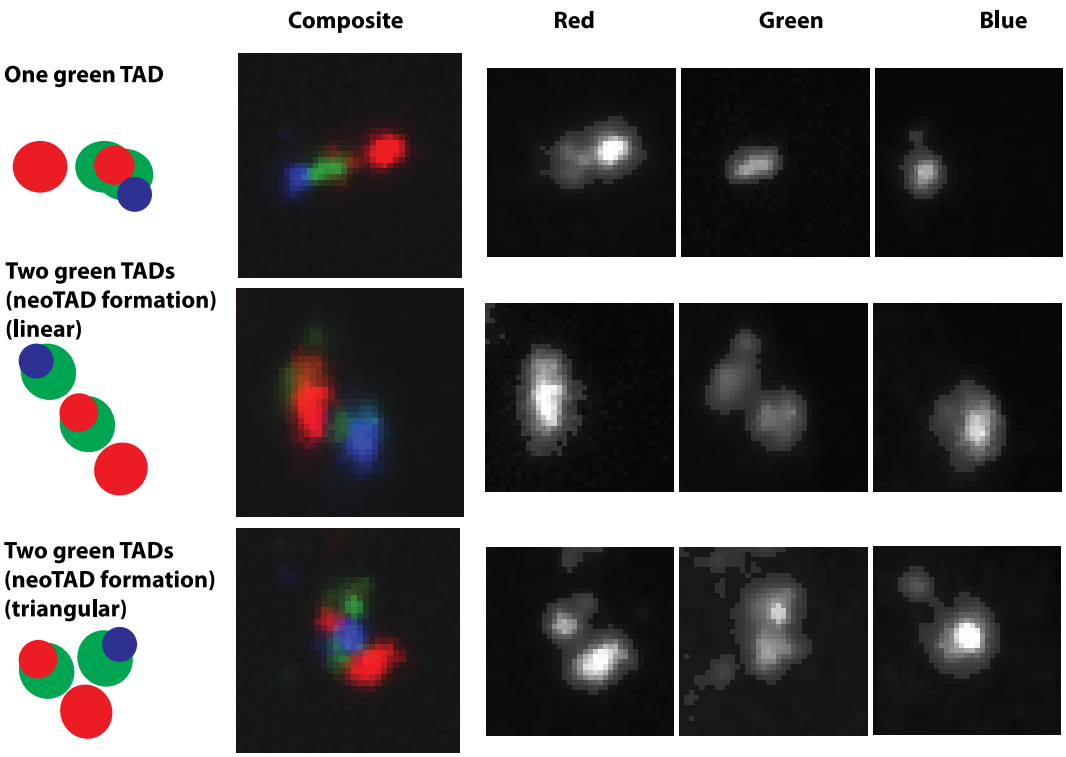
8% more red than blue in ESC
9% more green than blue in ESC

27% more red than blue in Dup
22% more green than blue in Dup

Figure 3.20: Comparison of cryoFISH areas in ES cells and Suox9 Duplication cells
A Dotplot of cryoFISH areas red, green and blue in ESC and Sox9 Duplication cells. For the volume analysis a selection of randomly chosen images from both replicates of ESC and Sox9 Duplication cells were analysed.
B The table contains the number of FISH signals taken into account for the volume analysis, their median area and respective standard deviation, as well as the median radius.
C Table shows the total number of red, green and blue FISH signals in the two ESC and Sox9 duplication cells replicates. Detection frequency of the blue probe is ~9% lower than the frequency of red and green FISH signals. In the duplication red and green signals are detected 27% and 22% (red, green) more frequently than the blue probe. Increased detection of green and red signals due to the duplication.

3.3.16 Visual inspection confirms, different TAD conformations are present at the Sox9 locus in Sox9 Duplication cells

In 3D image analysis it is possible that not all signals are detected as two distinct foci by the image analysis pipeline, because a per image global threshold can mistakenly join closely positioned but individual signals and count them as one signal. Visual inspection allows to quantitatively assess the locus configuration that FISH probes are engaged in. To avoid bias, visual inspection was performed by two people. The assessment of locus structure was focussed on the green signal, whether based on the signal spread in x/y/z it was perceived as one or two TADs. According to this classification, a large merged green signal implies the fusion of the duplicated domain with the Sox9 TAD, corresponding to a one TAD scenario. On the other hand, two separate green signals indicated a two TAD scenario, where the duplication was separated from the Sox9 TAD by formation of a *neo*TAD. Figure 3.21 shows examples of the two TADs and one TAD scenarios. The table in Figure 3.21 shows that over 38% of Sox9 duplication loci were classified as one TAD conformation, where the Sox9 TAD and the duplicated region create one large TAD. The two TAD conformation, where the duplicated region forms a separate neoTAD is assumed by 28% of the loci. 11% of the loci engage in a local structure where all labelled regions are overlapping to the same extent, which prohibits assignment of different TADs. The occurrence of 11% of all probes overlapping highlights the occurrence of subpopulations where different neighbouring TAD regions engage in preferential contacts across the TAD boundaries. A further 23% of regions could not be classified definitively. Unexpectedly, visual inspection revealed, that the 3D arrangement is varied and a “linear” TAD arrangement as well as a “triangular” TAD organization occurs (Figure 3.21). In the latter, the Sox9 TAD, which in the linear sequence is at a greater distance of the Kcnj TAD than the duplicated region, appears at an equivalent or shorter distance than the intervening duplicated region from the Kcnj TAD. A possible biological explanation for the triangular arrangement could be the intention to restore in part the original conformation that is present in the wild type mouse ES cells and thereby minimize the number of disrupted interactions between the neighbouring TADs.



Number of green TADs (conformation)	% of alleles
One green TAD	38%
Two green TADs (linear)	15%
Two green TAD (triangular)	13%
All overlapping	11%
Unclear	23%

Figure 3.21: Visual inspection of 3D FISH detects both the neoTAD and one TAD conformation
3D stacks of duplication cells were analysed by visual inspection, loci were classified according to the conformation of the green probe. 38% of loci were found to represent the one green TAD scenario and 28% of alleles were detected in a neoTAD conformation.

3.4 Discussion

3.4.1 Single cell methods show variability in TAD folding

Since their discovery, TADs have been regarded as one of the building blocks of chromosome folding. It has often been under debate to what extent TAD boundaries permit inter-TAD interactions. In particular, prevention of cross boundary interactions has been connected to the restriction of promoter-enhancer contacts (Lupianez et al., 2015; Anderson et al., 2014). The data presented here from cryoFISH and 3D FISH suggests a more subtle picture of TAD organization, where the organization of genomic regions within a TAD is highly variable on a single cell level. Most TADs undergo a stronger intra-TAD than inter-TAD interaction, however, FISH data shows a high degree of interaction between neighbouring TADs. In mouse ESC, 3D FISH shows that regions separated by the TAD boundary intermingle in more than 60% of cases with more than 20% of their volume. Also, 3D FISH in mouse ESC showed that 11% -14% of regions separated by the TAD boundary intermingle almost completely (more than 80%). This observation is in line with a recent report looking at inter-TAD interactions in single cell Hi-C (Flyamer et al., 2017). With 3D FISH, they showed that a pair of genomic regions, separated by a TAD boundary, was located at closer distance than an intra-TAD pair in 18% of cells, despite the inter-TAD pair having a twofold linear genomic separation.

Data obtained from 3D FISH and cryoFISH at the Sox9 locus therefore strengthens the view that TADs are population averaged events, arising from population-based methods such as Hi-C. TAD boundaries do affect median inter-probe distances but do not affect inter-TAD interactions on a single cell level. Also, population averaging does not consider the different conformations of TADs and neighbouring TADs on a single cell level. Possibly, also differential gene expression within the cell population can be attributed to variations in structure on a single cell level, a link recently suggested in a study by Giorgetti et al. (Giorgetti et al., 2014).

3.4.2 In Sox9 duplication cells widespread reorganization of the locus organization is occurring and a mixture of novel locus conformations arise

The analysis of center-of-mass distances obtained from 3D and cryoFISH data in homozygous ESC Sox9 Dup-L cells that contained a TAD boundary revealed wide ranging locus rearrangements. Upon insertion of the duplication the median distance between the previous inter-TAD regions green and red decreases, the median distances between the previous intra-TAD regions green and blue increases. The distance between the previous intra-TAD regions red and blue also increases. An increase in the blue green center-of-mass distance shows that the previous intra-TAD distance is overcompensated by a possible inter-TAD distance (neoTAD scenario) or in a one TAD scenario, the two green signals would move into such close spatial proximity, that they are detected as one large object, with a center-of-mass that is now located at greater distance from the green center-of-mass. Considering that the number of detected green objects was not doubled in 3D FISH and cryoFISH in Sox9 duplication cells as compared to blue signals, the predominant occurrence of the one TAD scenario explains the cryoFISH and 3D FISH results best. However, with 22% more green signals than blue signals in cryoFISH, the neoTAD scenario is occurring in a sub-population of cells. This result is also confirmed by visual inspection where ~28% of loci are observed forming a neoTAD and ~38% are engaged in a one TAD conformation. Only a small increase of 4% more green signals was detected in 3D FISH. A potential underestimation of the green signal in 3D FISH is possible, because of the lower resolution of 3D FISH and by adjusting the threshold per image, stack segmentation can potentially fail to discern two signals that are in very close proximity but nonetheless separate.

The Kcnj TAD itself was thought to not be affected by the duplication, hence a doubling of red signal was expected. However, 14% more red signals were detected in 3D FISH and 27% more in cryoFISH. This implies that a large fraction of duplicated red Kcnj regions locates in close proximity of the Kcnj TAD, where they cannot be detected as separate entities. By associating closely with its original genomic environment the duplication could possibly maintain more interactions that are present in the same region in wildtype mouse ESC. This finding also correlates with the observation that by visual inspection, two different neoTAD conformations were detected: a linear arrangement and a triangular arrangement of the two TADs. The triangular conformation places both the neoTAD and the Sox9 TAD at equal distance from the Kcnj TAD, thereby all three TADs can engage in inter-TAD contacts.

Based on cryoFISH and 3D FISH it can be concluded that two conformations arise in mESC Sox9 Duplication cells across the duplicated locus. The dominant conformation is one large TAD, where the original and the duplicated TADs are merged, the second arrangement is a neoTAD scenario, where the duplicated regions form a separate TAD. This study illustrated single cell heterogeneity regarding TAD conformations in a duplicated region that spans a TAD boundary.

3.4.3 Changes in volume after duplication in the Sox9 locus

Both in 3D FISH and in cryoFISH volume/area changes occur as a result of the duplication, however the measured changes are not equivalent with the two methods. In 3D FISH the duplication leads to a large scale decondensation of all three probes, whilst little additional signal is detected. In cryoFISH the blue and red probes become more compacted, and the green probe increases its volume. Also 22% additional green signals and 27% more red signals were detected. A decreased red median probe area can be partly explained by the 27% signal increase, which represents the small red duplication. In parallel many duplicated red regions co-associate with the red Kcnj TAD. The green probe increases its median area/volume both in 3D FISH and in cryoFISH, in the latter despite the 22% increase in detected green signals. Generally, an increase in green probe volume/area strongly suggests the formation of a single TAD, as a single TAD scenario would comprise twice the genomic content of the green probe and therefore, if volumes/areas remained constant, volumes would be expected to increase. Based on the evidence presented here, that predominantly one TAD forms and the neoTAD comprises a subpopulation, the green area increase is expected to arise due to the doubled genomic coverage.

3.4.4 Comparison of 3D FISH and cryoFISH

The experiments to test the neighbouring TAD configuration in mouse ESC as well as the rearrangements occurring in Sox9 duplication cells around the Sox9 locus have been performed both with 3D FISH and cryoFISH. The center-of-mass distances derived from both methods have shown the same general trend, while in absolute numbers the cryoFISH distances were consistently larger than the equivalent 3D FISH distances. In accordance with this, also the radial expansion of the Sox9 and Kcnj TAD FISH probes was greater than their respective radial expansion measured with cryoFISH. Spatial expansion in 3D FISH has also

previously been detected by Mongelard et al., where it was linked to effects occurring during the heat denaturation step in the 3D FISH protocol (Mongelard et al., 1999).
Regarding greater 3D center-of-mass distances, a technical explanation for this discrepancy could be the different fixations that are applied to cells that undergo 3D FISH versus cryoFISH. Fixation during cryoblock preparation is performed with more highly concentrated fixative and for a longer duration. It is possible that a shorter incubation period in a weaker fixative does not provide the same degree of conservation of the nuclear ultrastructure. Since the effective increase in distances and volumes is global, it is thinkable that the whole cell becomes enlarged as a result of the fixation and simultaneously induces an isotropic expansion of the chromatin. However, two studies have demonstrated the negligible effects 4% fixation has on cellular spatial dimensions: a) Li et al. showed that that a fixation 20 minutes in 4% PFA has only minor effects on cell size (Li et al., 2017) and b) Solovei at al. demonstrated that 10 minutes fixation in 4% PFA induces minor swelling/shrinking of the nucleus, however, small scale changes did not affect relative positions of labelled chromatin domains (Solovei et al., 2002). Cells treated with weaker fixation are however more susceptible to chromatin changes during the denaturation step in FISH, possibly evoking changes and partial decompaction of loci.

4 Super-enhancers engage in long-range interactions

4.1 Introduction

In the past years, research in the field of chromatin architecture has elucidated the formation of TADs and subTADs as basic features of chromatin organization (Dixon et al., 2012; Nora et al., 2012; Dowen et al., 2014). New findings have also shown that chromosomes are organized hierarchically and neighbouring TADs engage in preferential interactions (Fraser et al., 2015). On the largest scale, the occurrence of chromosome territories has been described (Cremer et al., 2001). However, until recently, the features of higher-order chromatin organization remained ill described. Whether for example long-range interactions between specific genomic regions occur at the length scale of several Mb and whether long-range gene-regulation via distal enhancers is a feature of 3D genome architecture. It has been proposed that enhancer-promoter interactions are restricted to the TAD level, more specifically that promoters and enhancers engage within cell-type specific subTADs (Javierre et al., 2016; Andrey et al., 2017). However, the use of genome architecture mapping (GAM) in mouse ESC revealed the occurrence of specific long-range contacts far beyond the TAD level (Beagrie et al., 2017). Using GAM, it was shown that a large number of long-range interactions occur between active-, enhancer- and intergenic-regions in mouse ESC. Long-range interactions span many tens of megabases, distances that were previously considered beyond the reach of specific interactions. For example, out of 4.5 million interactions involving active genes, 1.5 million span distances greater than 60 Mb (Beagrie et al., 2017). A further novelty detected by GAM is the detection of higher-order (multimer) contacts, i.e. the simultaneous interaction between three or more genomic regions. This information provides novel insight into 3D genome architecture and chromatin organization that occurs beyond the TAD level. In Hnisz et al., 2013, a new class of strong enhancers was discovered, so called super-enhancers, which have been attributed to regulate key cell identity genes (Hnisz et al., 2013). Analysis of GAM data found super-enhancers enriched in triplet contact formation, prompting us to investigate their involvement in long-range interactions and the occurrence of multimer contacts. Long-range interactions between regulatory elements and clustering of enhancers may provide a new understanding of genome organization and gene regulation.

4.2 Research motivation

In this work, I investigated long-range interactions between TADs containing super-enhancers at the single cell level, using cryoFISH and 3D FISH. I studied the formation of multimer contacts between highly interacting super-enhancers, probing the number of super-enhancer TADs involved in cluster formation. Performing FISH on triplets on two different chromosomes, I measured the distances between super-enhancers and whether they are more frequently colocalised with each other than with a control probe that is more closely located in the linear genome. Transcriptionally active regions tend to be more decondensed than lowly transcribed regions. To test whether super-enhancers are more decondensed than the lowly transcribed control region, I performed volume measurements of super-enhancer regions labelled with FISH probes in 3D FISH and cryoFISH. To test whether clustering changes with differentiation, I examined the formation of super-enhancer clusters in mouse ESC, terminally differentiated neurons and 3T3 fibroblasts. Finally I investigated whether long-range contacts between super-enhancers occur preferentially at specific nuclear landmarks.

Selection of probe positions in this work was done in collaboration with Robert Beagrie and Markus Schueler. Robert Beagrie, Markus Schueler and Antonio Scialdone performed the computational GAM analysis. Robert Beagrie has performed the computational analysis of the top and bottom super-enhancer regions. BACs were amplified and confirmed using PCR by Alexander Kukalev, and BAC probes were labelled by the Roukos lab (IMB Mainz). Sections of this work have been published in Beagrie et al. (2017).

4.3 Results

4.3.1 GAM detects multimer contacts between TADs containing super-enhancers

Chromosome conformation capture methods, such as Hi-C, preferentially detect pairwise contact frequencies between two genomic regions. Genome architecture mapping (GAM), on the other hand, can detect the occurrence of multivalent chromatin interactions, for example triplet interactions, which are contacts between three genomic regions that engage in a simultaneous contact. Figure 4.1A shows a hypothetical triplet contact, with three genomic regions colocalizing in the nucleus. In a typical two-way contact matrix, this event appears as

pairwise interactions of the three genomic regions. Based on the GAM matrix, it is however not possible to distinguish between a true triplet, a simultaneous colocalization of all three regions and a false triplet, where the three regions do not colocalize at the same time, but instead frequently interact in a pairwise fashion. To extract true triplet interactions, SLICE (StatisticaL Inference of Co-sEgregation) was applied. SLICE infers the probability of interaction between pairs of loci from GAM data by finding the proportion of estimated interacting and non-interacting co-segregation frequencies that best explain the observed co-segregation of that locus pair. Using SLICE, a triplet score can be calculated that reflects the likelihood of simultaneous triplet interactions for each possible combination of three TADs. This was calculated by first identifying sets of three TADs where all three TADs show high pairwise interaction frequencies. For those triplets a triplet interaction probability (Pi_3) was calculated. All triplets were ranked by their mean triplet Pi_3 value, and the top 5% were chosen for enrichment analysis. An exemplary GAM matrix that shows a triplet interaction is shown in Figure 4.1B, for a region that covers 35Mb in chromosome 1. Each pairwise interaction between TAD1, TAD2 and TAD3 shows a high probability of interaction in the GAM matrix; the enlarged pictures on the left show the specific interactions between the TADs. TADs were classified as super-enhancer (SE) TADs if they contained a super-enhancer from previously identified super-enhancers (Whyte et al., 2013). Non-SE TADs were classified based on their level of transcription as high (high), medium (medium) or low transcribed (low) using GRO-seq data from published data sets (Min et al., 2011). High, if they had GRO-seq coverage above the third quartile and low, if their coverage was below the first quartile. TADs between the upper and lower quartiles of coverage were classified as medium transcription. To select the TAD triplets most likely to interact simultaneously in the same cell, the highest scoring 2% (~101,000 triplets) were considered. Surprisingly, triplets occur most frequently amongst TADs that contain super-enhancers (Figure 4.1C). This type of preferential complex associations did not extend to TADs containing typical enhancers. Triplets are also frequently found between highly transcribed TADs or combinations of highly transcribed and super-enhancer TADs (Figure 4.1C). Based on this analysis, triplets between highly transcribing TADs occur second most frequent. Third and fourth ranking are triplet interactions between two highly transcribed TADs with one SE TAD and one highly transcribed TAD with two SE TADs, respectively. Triplet formation is depleted for interactions between three lowly transcribed TADs and combinations of lowly transcribed TADs with highly transcribed TADs or super-enhancer containing TADs.

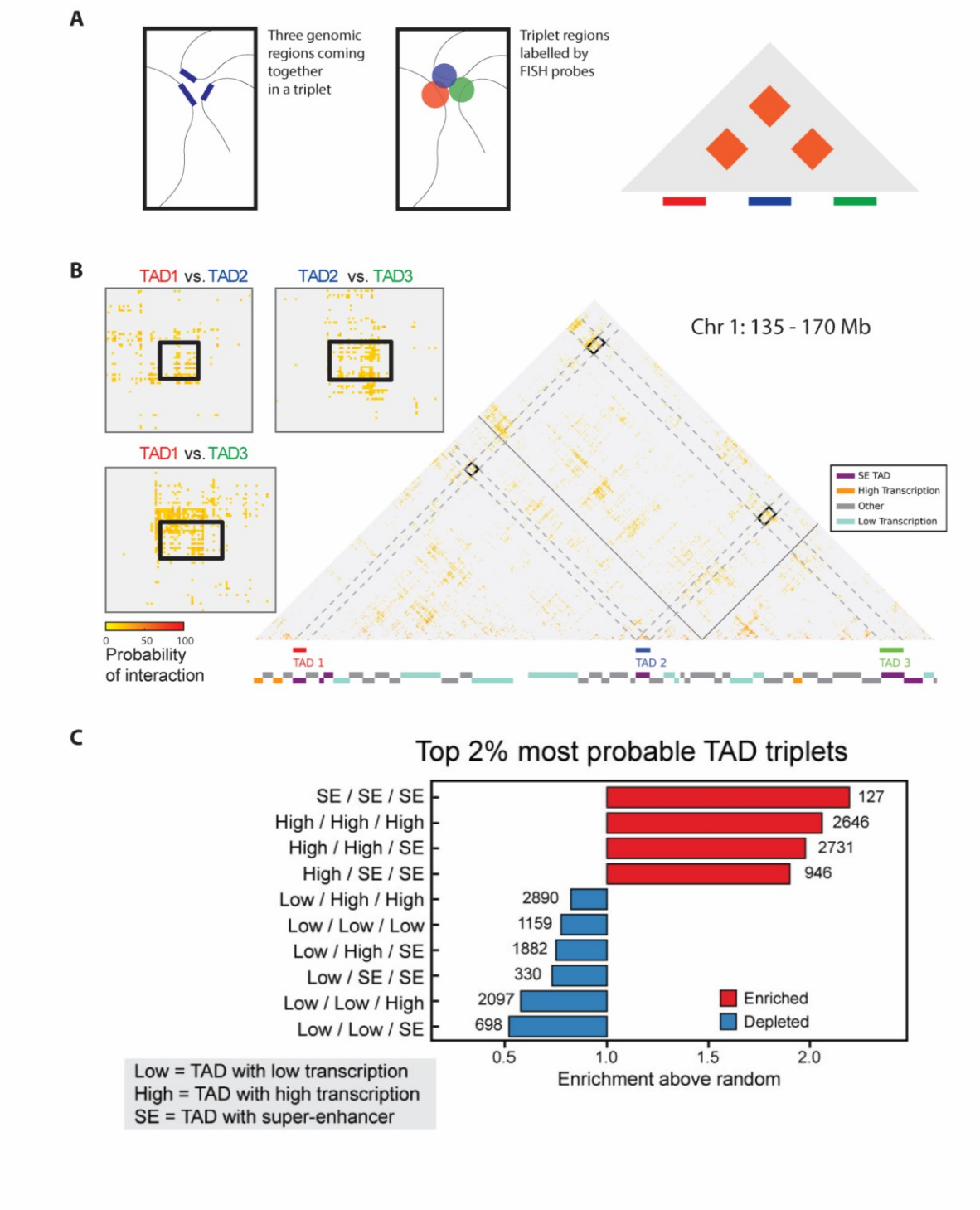


Figure 4.1: GAM detects formation of super-enhancer triplets

A The diagram represents a triplet interaction, which is a genomic configuration where three regions colocalize in close spatial proximity. In FISH, this would show in a high degree of colocalization of the three regions. In a GAM pairwise matrix, a triplet appears as the three regions red, green and blue engaging in mutual pairwise interactions. Implementation of SLICE allows distinction between true and false triplets.
B Example of a three-way interaction between TADs on chromosome 1 detected by SLICE. Large matrix shows prominent pairwise interactions over the entire region; small matrices show zoom of prominent interactions between the three TADs.
C Enrichment and depletion of different TAD classes in triplet interactions.
(Modified from Beagrie et al. 2017)

4.3.2 Super-enhancers form multimer interactions

To investigate long-range super-enhancer contacts and the formation of super-enhancer multimers, we analysed the cumulative frequency of triplet formation amongst the 2% highest scoring triplets (Figure 4.2A). Within this selection, the top 20% most triplet forming super-enhancer TADs are marked in red (top) and the 20% least often triplet forming super-enhancer TADs are marked in blue (bottom). Based on these two groups of super-enhancer containing TADs, we designed two FISH probes, the top and bottom super-enhancer probes, each covering 18 different super-enhancer TADs (Figure 4.2B). Each of the 18 labelled super-enhancer TADs per top and bottom probe set occupied a genomic region of 400 kb. The individual regions were designed by taking the TAD boundary closest to a super-enhancer within the set and extending it by 400kb. By including multiple super-enhancer TADs per top and bottom probe set, I could test whether clustering amongst the super-enhancer regions could be detected and whether clustering behaviour differed between the top and bottom probes.

From the total number of top 20% most interacting super-enhancer TADs and the total number of the 20% least interacting super-enhancer TADs, we choose the TADs to be included in the FISH probe sets according to two criteria: (1) a given chromosome should contain at least two top/bottom TADs, and (2) the distance between TADs had to be balanced between the top and bottom probes, to avoid a distance bias for one FISH probe set (Figure 4.2C). The probe distribution over the chromosomes is shown in Figure 4.2B, indicating the top most interacting super-enhancer TADs in green and least interacting super-enhancer TADs in red. All chromosomes, except chromosome 3, contain probes of only one colour. Chromosome 3 contained two bottom super-enhancer and three top super-enhancer TADs. The distances between the chosen super-enhancer TADs within each varied between 7 Mb and 97 Mb (Figure 4.2C). Figure 4.2D shows that the number of normal enhancers within the 18 regions of the top and bottom probes was also similar between the two probe sets. The transcriptional activity within the two probe sets is different and the mean transcription within the top probes is slightly higher than within the bottom probes (Figure 4.2E). GRO-seq data are obtained from published datasets (Min et al., 2011). FISH probes were custom designed by myTags (Arbor Biosciences) as oligonucleotide probes directly conjugated with fluorochromes. Top super-enhancers were labelled with ATTO550 and bottom super-enhancers with ATTO594.

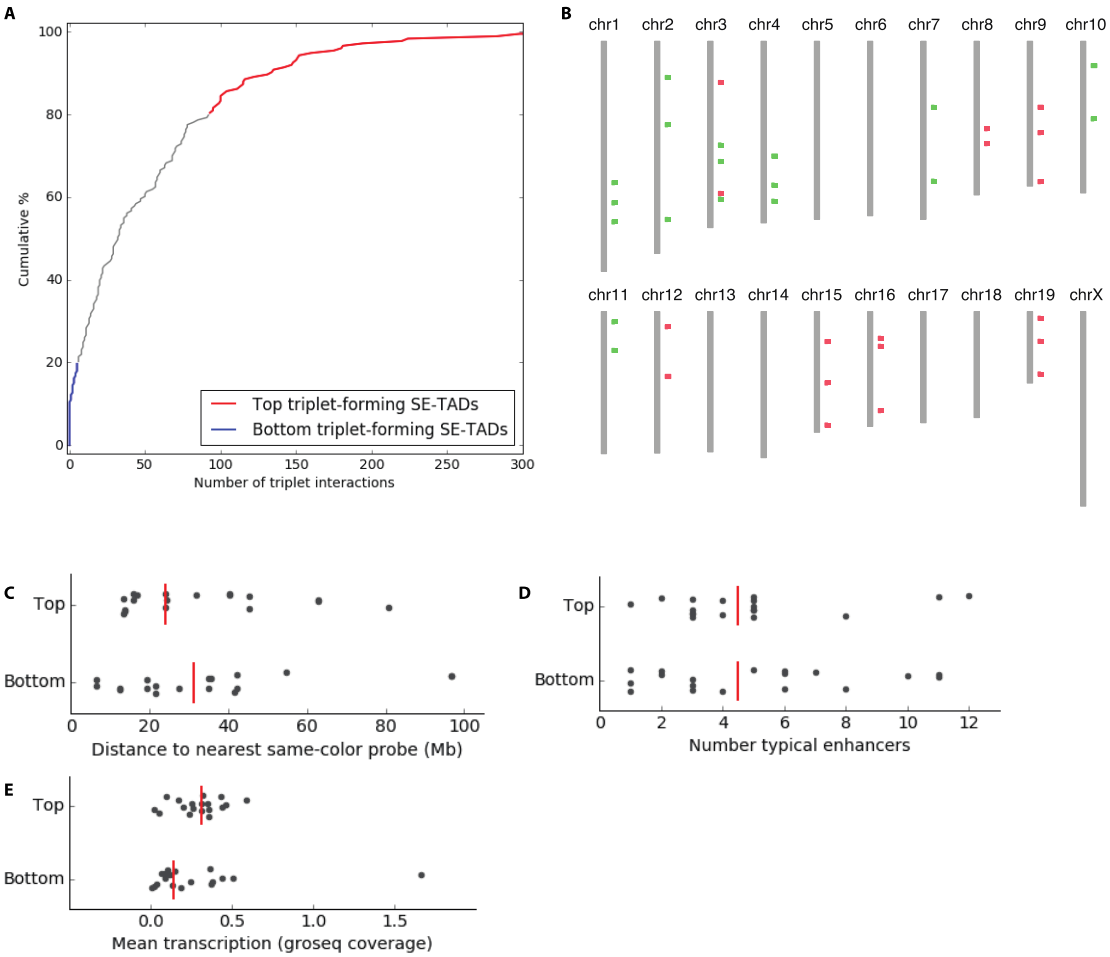


Figure 4.2: Investigating the spatial organization of super-enhancers that are ranked in the 20% most often found in triplets and the 20% bottom triplets

A Cumulative frequency of the super-enhancers engaged in triplet. Super-enhancers that are ranked as the 20% most highly triplet forming are shown in red and the 20% bottom super-enhancers found in few triplets are marked in blue.

B Distribution of the probe regions within the top probe and the bottom probe. Regions labelled in green are part of the top probe set and regions labelled in red are part of the bottom probe set.

C The distances to the nearest colour probe within the set of top and bottom enhancers is shown. To avoid a proximity bias between the probes of one set over the other set, probe selection was based on making average inter-probe distances as equal as possible.

D The number of typical enhancers contained within the individual probe regions.

E The mean transcription (GRO-seq coverage) of the top and bottom probes (GRO-seq data from Min et al., 2011). Top probes show a slightly higher mean transcription than the bottom probes.

These analyses were performed by R. A. Beagrie.

Using the top and bottom probes, I first tested whether super-enhancers TADs form clusters in mESC, as indicated by GAM. I also investigated whether the clustering behaviour of the top super-enhancer TADs is different from bottom super-enhancer TADs. Different possibilities were conceivable:

1) Top and bottom probes could differ in their cluster sizes, with top probes clustering more extensively, and forming potentially larger clusters. Bottom super-enhancers could cluster to a lesser extent, presenting more isolated positions, with fewer contacts to other super-enhancer TADs. In FISH this would results in a difference in volume/area size between the top and bottom probes.

2) Alternatively, the observed clustering between top and bottom probes could be similar, because top and bottom super-enhancers are dissimilar in the number of different clustering partners they engage with. For example, bottom probes could have more stable interactions with fewer super-enhancers, whilst top probes might engage in interactions with many different SE TADs, but the total sum of interaction partner might be similar. In FISH this could show in similar sized signal volumes/areas of the top and bottom super-enhancers.

I started by performing 3D FISH simultaneously with the top and bottom super-enhancer probes (ATTO550 and ATTO594, respectively) in mouse ESCs. I detected a striking degree of cluster formation by visual inspection of the 3D z-stacks. Figure 4.3 shows two images of single optical sections taken from two independent z-stacks. The top probes are shown in green and the bottom probes in red. The magnified images in Figure 4.3 show zoomed in examples of clustering, with several FISH signals arranged in a linear and clumped way. Equally, individual FISH probe signals representing unclustered super-enhancers were detected. The images in Figure 4.3 also show evidence that clustering is not limited to triplets, for example in panels 1, 2, 4 clusters of up to five super-enhancer TADs are seen. Interestingly, I observed many contacts between top (green) and bottom (red) probes, as shown in Figure 4.3 panels 1, 2, 4. As only chromosome 3 contained top and bottom probes and all chromosomes were covered by a maximum of three probes of each colour, clusters of more than three super-enhancers of one colour as well as very frequent contacts between top and bottom super-enhancers, suggests that clusters of super-enhancer TADs also form in trans, between chromosomes.

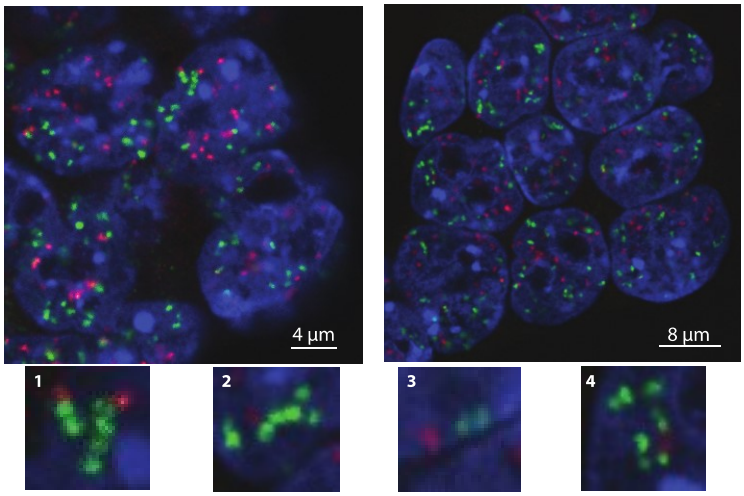
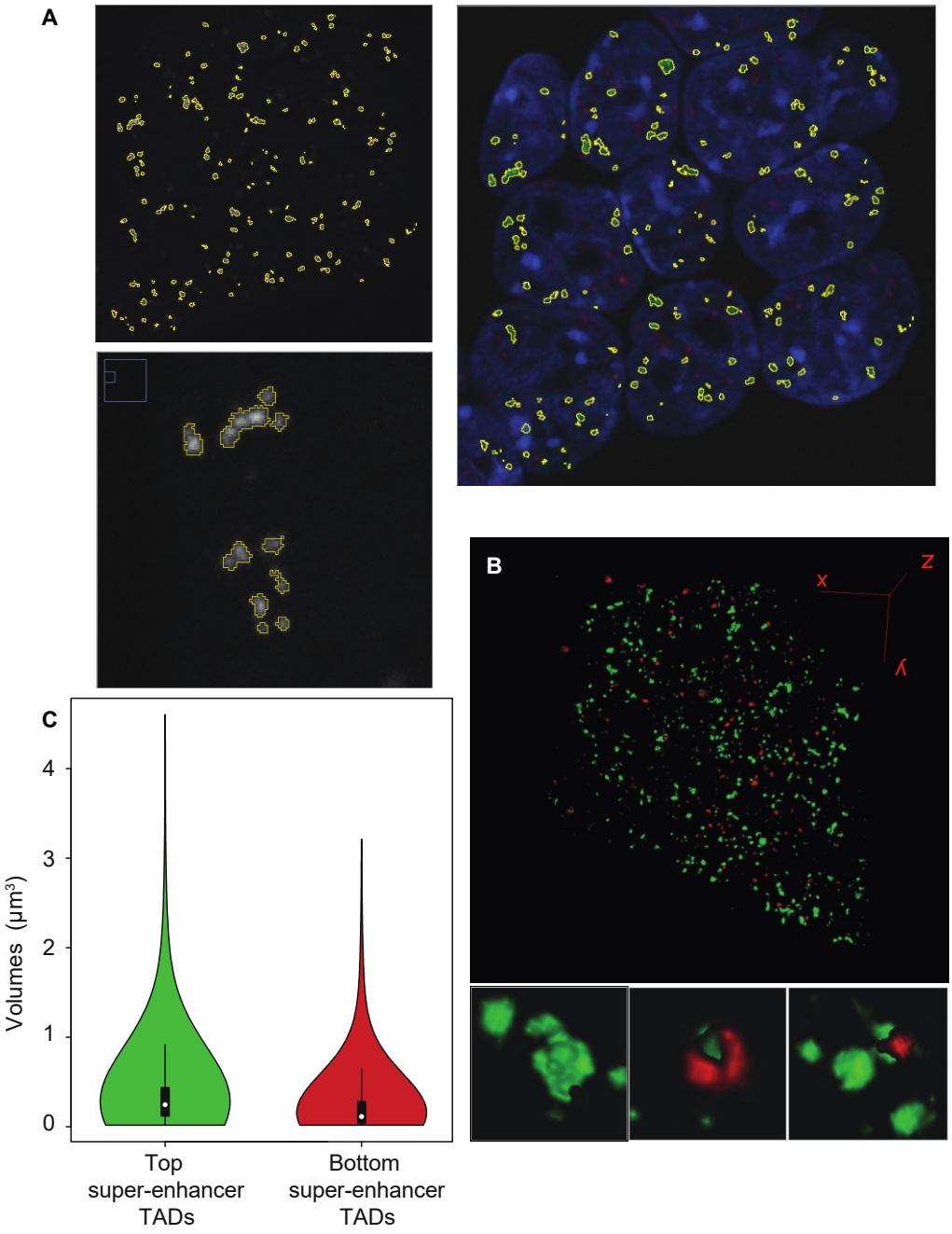


Figure 4.3: Extensive clustering is seen in 3D cells labelled with FISH probes covering the 20% most interacting and the 20% bottom interacting super-enhancers
ESC nuclei are shown in blue, top interacting super-enhancers are green and bottom interacting super-enhancers are red. Both images A and B are optical sections taken from 3D z-stacks. The magnified images show examples of extensive probe clustering and examples of individual probes not located in a cluster. Furthermore clustering between top (green) and bottom (red) probes is detected.

The clustering behaviour of super-enhancer TADs can be quantified, for example by exploring the volume of contiguous signals. Both highly clustered and individual probe signals were observed by visual inspection. To quantify the number and 3D volumes of the top and bottom super-enhancer FISH probes and calculate the number of interaction partners within the clusters, I analysed images of three z-stacks of mouse ESC cells, labelled with top and bottom super-enhancers. First, image segmentation for each channel was done in Image J. Thresholding of the channels was done separately for each colour and was based on visual inspection. An exemplary segmentation of a raw optical section, taken from a stack is shown in Figure 4.4A. To obtain the probe volumes and the total number of signals, segmented images were subsequently analysed in ImageJ by applying the ImageJ 3D particle analysis tool and visualized using 3D reconstruction. To exclude background signal from individual voxels, a consistent size filter for both bottom and top probes was implemented. Figure 4.4B shows the 3D signal reconstruction of a whole 3D stack, nuclear staining is not shown. 3D reconstruction revealed interesting configurations formed by the clusters. Panel 4.4B shows magnified examples that highlight the formation of large clusters of super-enhancer regions.

The distribution of the volumes of top (green) and bottom (red) regions is shown in Figure 4.4C as a violin plot. Top probes have a higher median volume and standard deviation than bottom super-enhancers (table Figure 4.4). Top probes occupy a median volume of 0.247 μm^3 and the median bottom super-enhancer volume is 0.114 μm^3, therefore the median volume of the top super-enhancers is twice the median volume of the bottom super-enhancers. The difference in radii between clusters is 23%. Radii were calculated on assumed sphericity of the clusters, but as super-enhancer clusters often occur in a linear cluster conformation, the cluster volume is likely to be a better indicator for cluster size.

The bottom and top super-enhancer TAD FISH signals were detected in a similar number of clusters (1395 and 1226, respectively). The difference in detection is about 12% and may reflect a tendency of the bottom super-enhancer TADs to be more often detected outside a cluster, versus a tendency of the top super-enhancer TADs to form larger clusters.



	Number	Median Volume (µm³)	Mean Volume ± Standard dev (µm³)	Median Radius (µm)
Top super-enhancers	1226	0.247	0.380 ± 0.457	0.389
Bottom super-enhancers	1395	0.114	0.230 ± 0.355	0.301

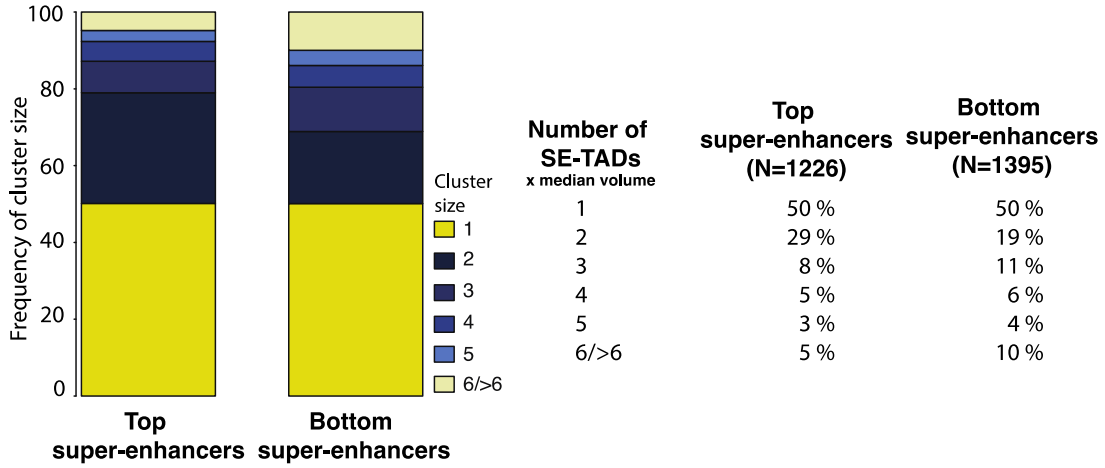
Figure 4.4: Super-enhancer clustering in 3D
A Steps of the image segmentation. The image on the upper left is an optical section of the green channel (top super-enhancers) from a 3D stack. The yellow selection is the signal contained in the thresholded images. The image on the bottom left is zoomed in and exemplifies the signal selection. The image on the right shows the signal selection for the top super-enhancers in the merged image. Top super-enhancers are shown in green, bottom super-enhancers in red, DAPI nuclear counterstain is shown in blue and the thresholded signal is shown in yellow.
B Volume comparison of top super-enhancers and bottom super-enhancers. The median volume of the top super-enhancers is twice as large as the median volume of the bottom super-enhancers. Also the standard deviation is higher in top super-enhancer clusters. For volume analysis three z-stacks of containing ~20 cell per stack, were analyzed. In total 1226 signals for the top super-enhancers and 1395 signals for the bottom super-enhancers were detected.
C The image shows a 3D reconstruction of the top super-enhancers (green) and bottom super-enhancers (red) in ESC from a z-Stack. The panels below show some examples of clustering of top and bottom super-enhancers .

The table contains the total number of FISH signals detected by image analysis, the median volume and respective standard deviation, as well as the median cluster radius (assuming sphericity).

To explore the degree of clustering and the number of super-enhancer TADs involved in the clusters, I used the approximation that the median volume of the top/bottom probes represents the volume of a single super-enhancer. The top super-enhancer radius (0.389 μm) compares well with the radius obtained for a single super-enhancers SE5 (0.386 μm) and is lower than the radius of a second single super-enhancer SE4 (0.547 μm), both measured in 3D (see Figure 4.15). In this scenario multiples of the median are clusters of multiple super-enhancer TADs. This is a simplified approximation, but it is in line with the observation that besides the larger clusters also isolated signals are detected, it might however underestimate cluster formation as visual inspection showed a greater tendency for probes to be inside clusters than isolated.

This approximation also does not take into account that in total more bottom probes were detected than top probes. By binning the measured volumes based on the multiplicity of the median volume, an approximation of the cluster sizes is obtained. Figure 4.5 shows the frequency of cluster sizes as multiples of the median. Both the bottom and top probes undergo extensive clustering, with many top and bottom probes engaging in clusters of three or more super-enhancers. 29% of the top super-enhancer TADs and 19% of the bottom super-enhancer TADs engage in pairwise contacts. In total more than 20% of the top super-enhancer TADs and over 30% of the bottom probes are in clusters of three or more interaction partners. Amongst this even 4% of the top and 10% of the bottom super-enhancers are in clusters of six or more TADs.

Based on this analysis, both top and bottom super-enhancers form large clusters containing distally located super-enhancer TADs. Bottom super-enhancers are even more frequently clustered than the top super-enhancers. Top and bottom super-enhancers could differ in their potential regulatory activity, as bottom super-enhancers occupy smaller volumes than top super-enhancers and typically more open chromatin regions are more active, which is also reflected by the GRO-seq coverage of the top and bottom super-enhancers (Figure 4.2E). Extensive clustering also invites the interpretation that the detected signals might be subsets of large clusters, that contain more than the labelled genomic regions.

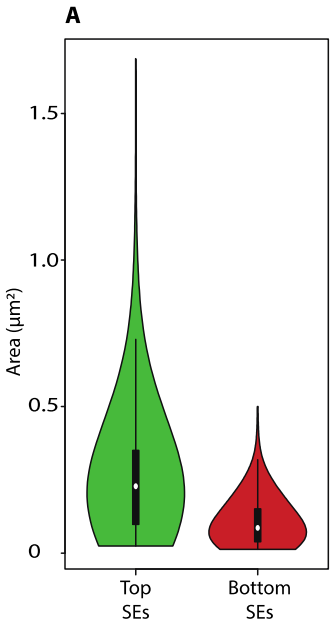


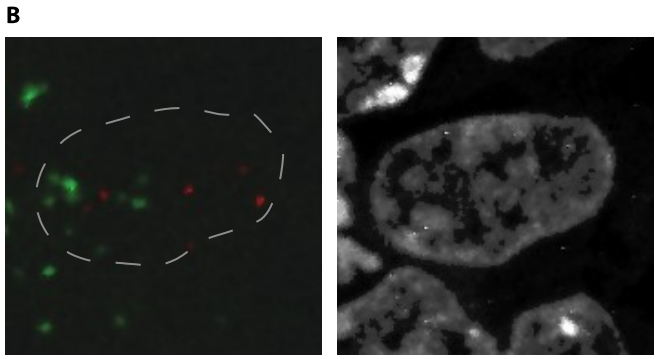
Number of SE-TADs x median volume	Top super-enhancers (N=1226)	Bottom super-enhancers (N=1395)
1	50 %	50 %
2	29 %	19 %
3	8 %	11 %
4	5 %	6 %
5	3 %	4 %
6/>6	5 %	10 %

Figure 4.5: Number of top and bottom super-enhancer TADs involved in the clusters
Assuming the median volume (Figure 4.4) represents the median size of an individual top/bottom probe, the number of super-enhancer TADs within clusters were calculated based on multiples of the median. Clustering is found to be extensive with more than 20% of top super-enhancers and ~30% of bottom super-enhancers engaging in clusters of three or more super-enhancer TADs. Due to the smaller median area, bottom super-enhancer probes were found to engage in larger clusters.

To achieve higher spatial resolution in the detected of super-enhancer clusters, and assess whether the same clustering behaviour can be observed in cryosections, I performed FISH using top and bottom super-enhancer probes on ~200 nm cryosections from mouse ES cells. Subsequently, I analysed the areas that top and bottom super-enhancer probes occupy in sections. Area analysis was performed using ImageJ, applying segmentation of signal in ImageJ and subsequently using the ImageJ particle analysis tool on the segmented areas, to extract area sizes. In Figure 4.6, the area distribution of the top and bottom probes in cryosections is shown. The violin plot in Figure 4.6A shows that the median areas occupied

by the top probes are again larger with a bigger standard deviation. The median area of the top probes is 3.35-fold greater than the median area of the bottom probes, 0.288 compared to 0.086 μm^2 (top and bottom, respectively). Figure 4.6B shows a representative image of top and bottom super-enhancer FISH signals. In 3D, 12% more bottom probe signals were detected, in cryoFISH this was reversed and instead 24% more top than bottom probe signals were detected. An important aspect that needs to be considered with respect to measuring clustering in cryosections is the increased imaging resolution which allows for distinction of independent objects even in situations where they might be clustered (at short physical distances).





	Number	Median Area (µm²)	Mean Area ± Standard dev (µm²)	Median Radius (µm)
ESC cryo - top	899	0.228	0.254 ± 0.201	0.269
ESC cryo - bottom	683	0.086	0.105 ± 0.080	0.165

Figure 4.6: Top super-enhancers occupy larger areas in ESC cryosections than bottom super-enhancers

A Shows the violin plots of the top and bottom super-enhancers measured in ESC cryoFISH. The median area of the top super-enhancer clusters is almost three times higher than the median area of the bottom super-enhancers.

B The image on the left is an example of top (green) and bottom (red) super-enhancers in ESC stained with DAPI (grey), the grey line represents the outlines of the DAPI counterstain. The image on the right shows the DAPI counterstain in grey.

The table contains the number of signals analysed, their median area, the standard deviation and the median cluster radius (assuming sphericity).

4.3.3 Long-range interactions occur between distal super-enhancer TADs

Using the top and bottom super-enhancer probes, I could detect cluster formation between distal super-enhancers in 3D ES cells and in mESC cryosections. However, due to the nature of the top and bottom FISH probes, containing 18 TADs per colour, it is not possible to retrieve information about the distance between individual super-enhancers and their frequency of colocalization. To investigate how often individual super-enhancer TADs come into close spatial proximity and at what median distances they are localized, we chose two super-enhancer triplets with a high triplet interaction score in GAM (Figure 4.7). Triplet 1 is located on chromosome 3, with the three super-enhancer regions separated by 7.9 Mb (SE1/SE2), 5.8 Mb (SE2/SE3) and 14.1 Mb (SE1/SE3). Triplet 2 is located on chromosome 11, and the super-enhancer TADs were separated by 9.7 Mb (SE5/SE6) up to 28.1 Mb (SE4/SE6). Two lowly transcribed TADs were chosen as control probes, positioned between SE1/SE2 and SE4/SE5, with a distance of 2.6 Mb (SE1/low) and 7.6 Mb (SE4/low)/ 10.2 Mb (SE5/low). GAM pi matrices in Figure 4.7 show regions encompassing two super-enhancer triplets on chromosome 3 (left) and chromosome 11 (right). The regions corresponding to the point of interaction between the super-enhancer probes and the low probes are highlighted by the boxes in the GAM matrices.

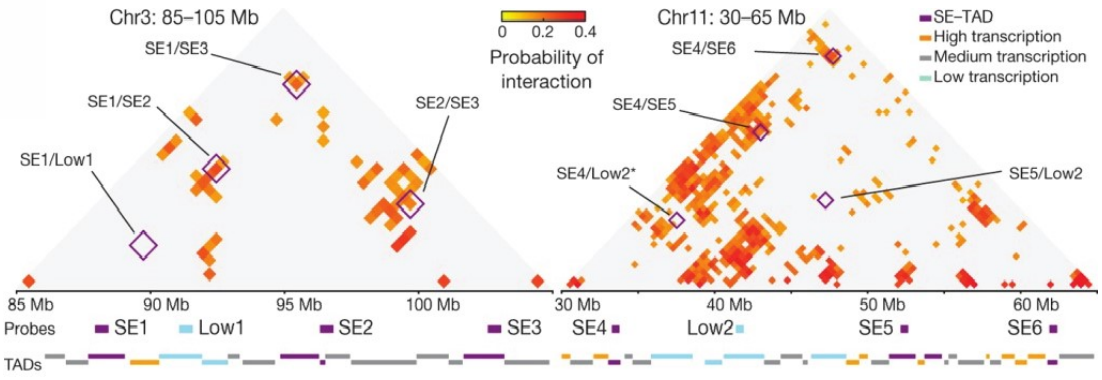


Figure 4.7: Super-enhancer triplets on chromosome 3 and chromosome 11
GAM Pi matrices of two triplets on chromosome 3 and chromosome 11. SE TAD regions chosen for FISH are highlighted in purple. TADs within the regions and their classification, are indicated at the left side; Yellow TADs are high in transcription, grey TADs are medium in transcription and TADs in light blue are lowly transcribed. Contacts between two super-enhancers are highlighted with black boxes.

To measure the distances between the super-enhancers, we obtained custom-designed myTag probes covering 500 kb of the super-enhancer TADs of triplet 1 and triplet 2 and the respective control probes on chromosomes 3 and 11. Fluorophores utilized to label the super-enhancer probes are shown in the table below, Table 4.1.

Table 4. 1 - Table 4.1 Fluoroescent myTag probes utilized for FISH experiments

Probe Name	Probe co-ordinates (mm9)	Labels used	Origin
SE1	chr3:87930000-88430000	ATTO550; ATTO594	Mytags
SE2	chr3:96320000-96820000	ATTO647N	Mytags
SE3	chr3:102580000-103080000	ATTO594	Mytags
SE4	chr11:33280000-33780000	ATTO550; ATTO594	Mytags
SE5	chr11:52150000-52650000	ATTO550; ATTO594	Mytags
SE6	chr11:61880000-62380000	ATTO647N	Mytags
Low1	chr3:91070000-91570000	ATTO488, ATTO647N	Mytags
Low2	chr11:41400000-41900000	ATTO488	Mytags

Figures 4.8 and 4.9 demonstrate the genomic regions covered by the FISH probes. Figures contain the position of the super-enhancers and genes within the region, as well as the H3K27ac signal and gene expression as shown by mRNA-seq (datasets were taken from ENCODE project in E14 mESC; Ferrai et al., 2017, respectively). Super-enhancer positions are based on Whyte et al. 2013. Super-enhancer FISH probes cover gene rich regions, the low probe regions are gene sparce and lowly transcribed.

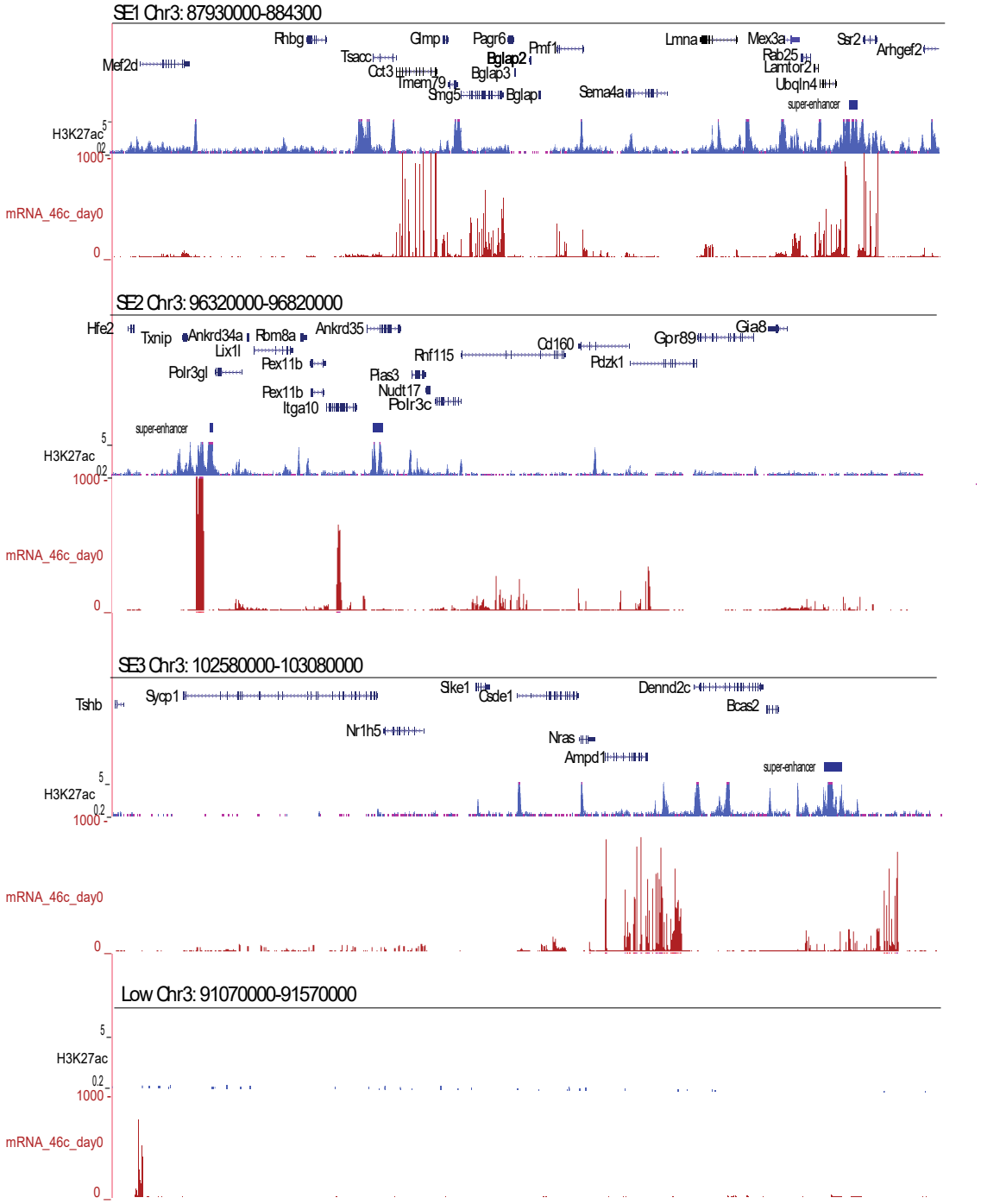


Figure 4.8: Position of the super-enhancer probes and control probes on chr 3
UCSC profiles are shown, with the genes within the probe region, the super-enhancer location, H3K27acetylation (from ENCODE project in E14 mESC) and gene expression based on mRNA-seq (from Ferrai et al., 2017).

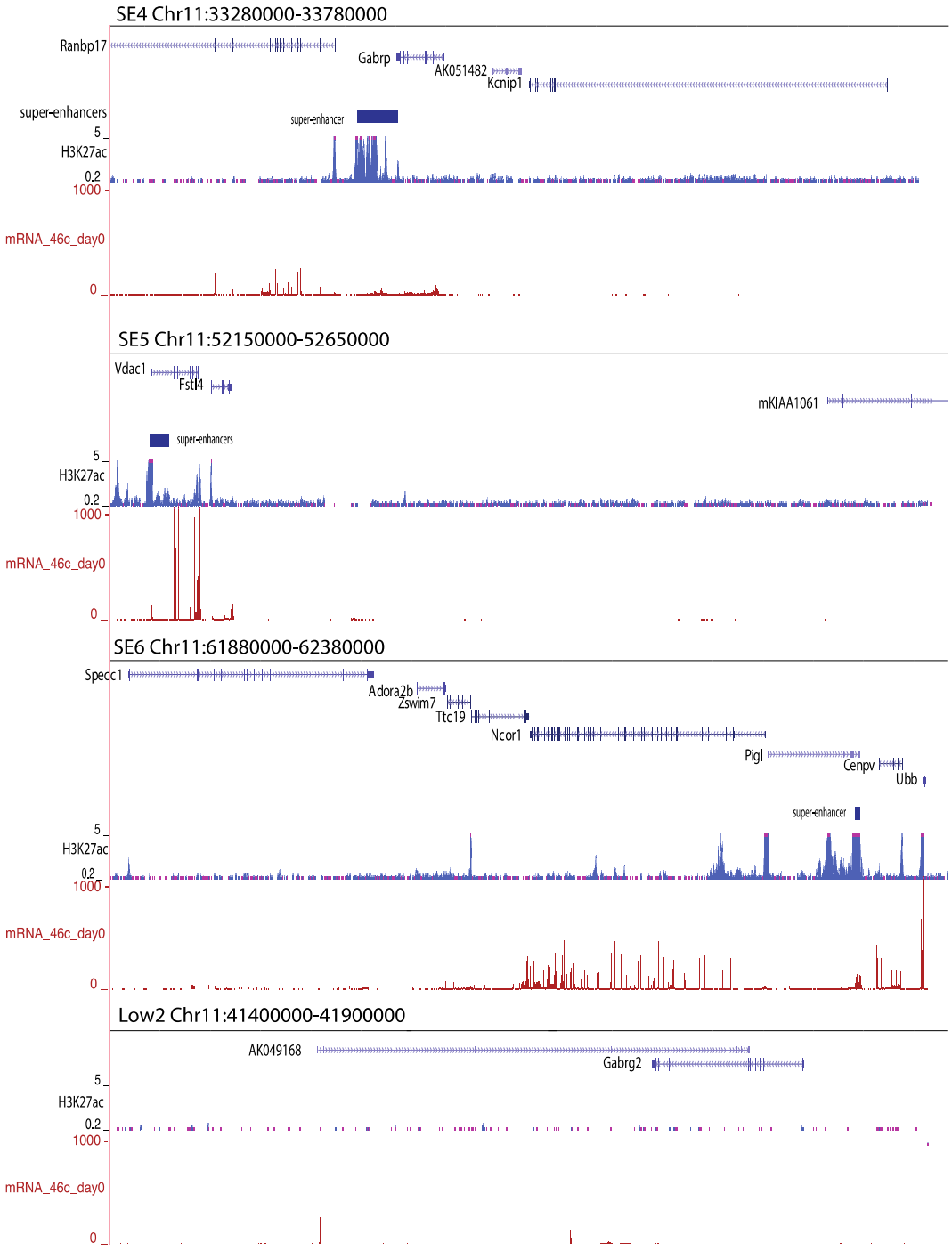


Figure 4.9: Position of the super-enhancer probes and control probes on chr 11
UCSC profiles are shown, with the genes within the probe region, the super-enhancer location, H3K27acetylation (from ENCODE project in E14 mESC) and gene expression based on mRNA-seq (from Ferrai et al., 2017).

The table in Figure 4.10A summarizes the respective GAM normalized linkage (normalized for detection frequency of individual loci) and interaction probabilies (as determined by SLICE). GAM normalized linkage and interaction frequencies are high for the super-enhancers within a triplet and low for the control probes. For example super-enhancers 4 and 5 are at 18.4 Mb distance, with 0.42 normalized linkage and 0.23 probability of interaction, whilst SE5 and low probe 2 are at a distance of 10.2 Mb with an interaction probability of 0.00 and a GAM normalized linkage of 0.21.

Co-hybridization of super-enhancer probes and control probes were performed on ~200 nm cryosections. Figure 4.10 shows example images of colocalized super-enhancers within the cell nucleus. For example, the upper right image in Figure 4.10A shows the super-enhancers SE1/SE3 (green and red, respecitvely) colocalized and the lower right panel shows the SE4/SE5 probes (green and red respectively) touching at the periphery. The upper panel on the left shows the absence of contact between SE5 and the control probe (green and red, respectively) and the lower panel on the left shows a SE4 and control probe 2 (green and gred respectively) in close proximity, but not overlapping. Figure 4.10B shows two examples of the triplet on chromosome 3, where all three super-enhancers are contacting each other. Some examples of triplets were detected but the frequency of triplet occurence was not measured.

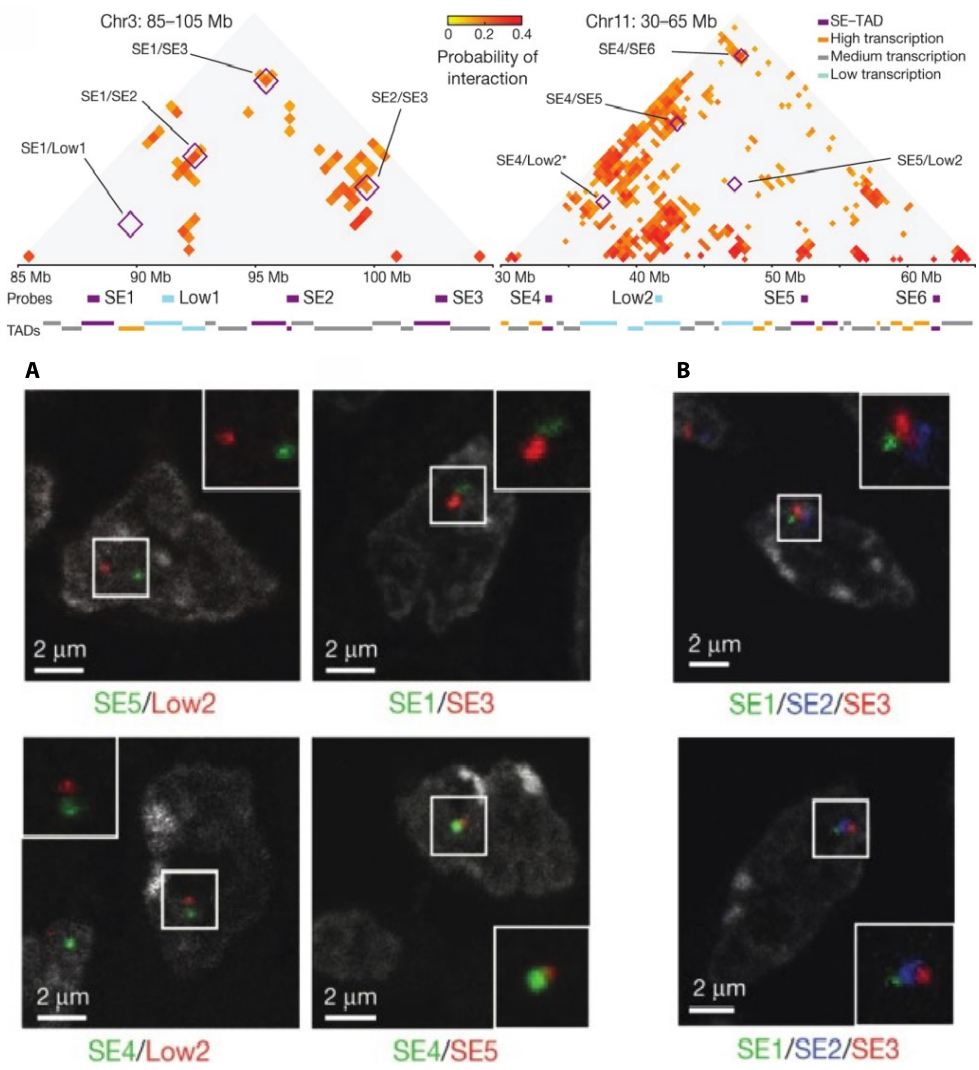


Figure 4.10: Example images showing long-distance super-enhancer contacts
A Example images showing super-enhancers and control probes. For example strong colocalization between SE1/SE3 and SE4/SE5. DAPI nuclear counterstain is shown in grey.
B Some example for triplet formation between SE1/SE2/SE3. DAPI nuclear counterstain is shown in grey.

By visual inspection FISH signals showed high degrees of decompaction prompting us to utilize edge-to-edge distance detection to measure inter-probe distances, instead of center-of-mass distances. The edge-to-edge distances between super-enhancer FISH signals were calculated on a Python-based edge detection script, which computed the closest distances between locus peripheries of different colour in the same nucleus based on segmented images obtained by an Image J macro, previsouly written by Miguel Branco (Ana Pombo laboratory, edge-to-edge script written by Robert Beagrie). In Figure 4.11 the edge-to-edge distances between the super-enhancers and the control probes are shown. Distances were calculated from a minimum of two replicates for each pair of probes. The grey line shows the expected distances based on the super-enhancer/control probe distances. All super-enhancer TADs are

at a closer median distance and more often colocalized at zero distance than they are to the control probe, despite the smaller inter-probe distance between SE1/Low1. Remarkably, super-enhancers 4 and 6 located at 28.1 Mb distance show a high frequeny of colocalization at 0 nm of 18% and super-enhancers 4 and 5, which are at a distance of 18.4 Mb, are colocalized at 0 nm in 74% of the measurements. Figure 4.11 Table B contains the median inter-probe distances as measured by edge-to-edge detection, as well as the average edge-to-edge distance and standard deviation. Median distances between the respective control probes and super-enhancers 1 and 5 was 777 nm (2.6 Mb) and 686 nm (10.2 Mb), which is higher than the median super-enhancer distances, which were located between 0 nm (SE4/SE5) and 681 nm (SE4/SE6). Surprisingly, super-enhancer SE4 and the low probe 2 shared an interaction probability of 0.12, which is reflected in the 26% frequency of colocalization at 0 nm.

The high frequencies of colocalization and the closer than expected median distances show that distal super-enhancers contact each other. Long-range interactions between regulatory elements of this length scale have not been previously reported and may be involved with novel long-range gene regulatory mechanisms.

Table A

	Predicted non-interacting probe pairs		Predicted interacting probe pairs (increasing genomic distance)					
	SE1 / Low1	SE5 / Low2	SE2 / SE3	SE4 / Low2*	SE1 / SE2	SE1 / SE3	SE4 / SE5	SE4 / SE6
Genomic distance (Mb)	2.6	10.2	5.8	7.6	7.9	14.1	18.4	28.1
GAM normalized linkage	0.13	0.21	0.40	0.38	0.41	0.29	0.42	0.38
Probability of interaction	0.00	0.00	**0.18**	0.12*	**0.21**	**0.13**	**0.23**	**0.23**

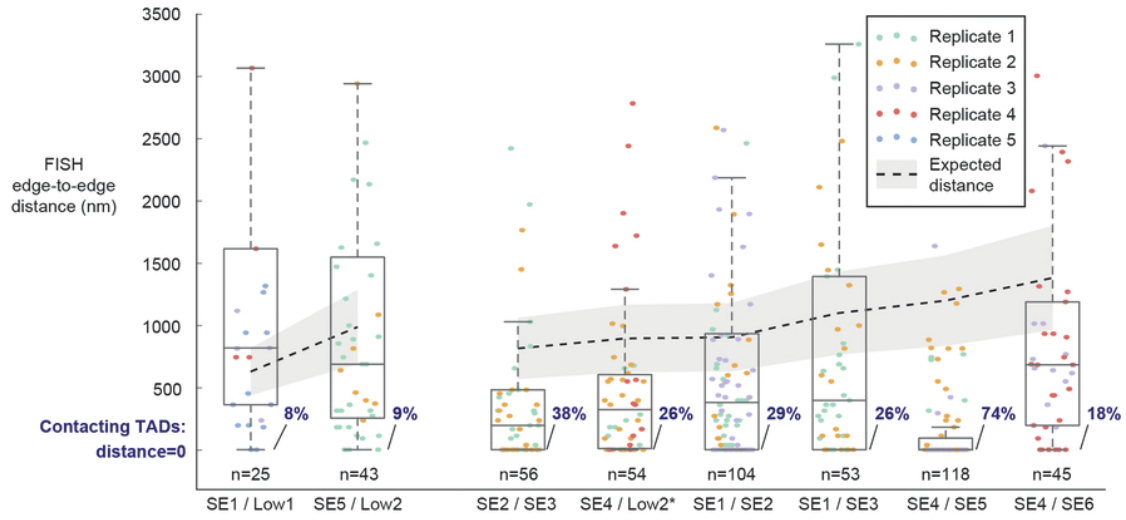


Table B

	SE1/Low1	SE5/Low2	SE2/SE3	SE4/Low2	SE1/SE2	SE1/SE3	SE4/SE5	SE4/SE6
Median edge-to-edge distance (nm)	777	686.	195	322	378.51	396	0	681
Mean edge to edge distance (nm) ± standard deviation	1842 ± 2441	1515 ± 2217	890 ± 2010	556 ± 926	1146 ± 2143	1214 ± 1992	432 ± 1311	1138 ± 1540

Figure 4.11: Super-enhancers are more closely located than to the control probe
The boxplot shows the edge-to-edge distances between the super-enhancers and the low probes. Super-enhancers are more often coloclized and located at a smaller median distance than to the low probe.

Table A shows the genomic distance (Mb), the GAM normalized linkage and the probability of interaction between the super-enhancers and the control probe.
Table B shows the edge-to-edge and center-of-mass distances between the FISH probes.

4.3.4 Super-Enhancers engage in long-range interactions

With experiments using FISH probes that cover 18 top and bottom super-enhancer TADs I could show that super-enhancer TADs contact each other form multimers. Using, 500 kb FISH probes that cover super-enhancer TADs, I could furthermore show that pairwise colocalization of super-enhancer TADs occurs with high frequency and close spatial distances. To investigate whether the super-enhancers are required to mediate the long-range contact, we designed BAC probes of ~180-200kb size covering the super-enhancers on chromosome 11, Figure 4.12. The BAC probes were chosen from a library of commercially available BAC probes (obtained from BACPAC //bacpac.chori.org/) in a way to maximize overlap with the previously used, 500 kb myTags FISH probes. Figure 4.12 shows the relative position of the BAC probes to the myTags probes and the super-enhancer position. Due to the smaller length of the BAC probes, the distance between the BAC super-enhancer FISH probes, as measured from the center of the probe increased by 400 kb (SE4/SE5) and 500 kb (SE5/SE6). BAC probe labelling was performed in Vassilis Roukos laboratory (IMB Mainz), fluorophores utilized for probe labelling were: SE4: AlexaFluor488, SE5: AlexaFluor 488, AlexaFluor 568 and AlexaFluor647, SE6: AlexaFluor568.

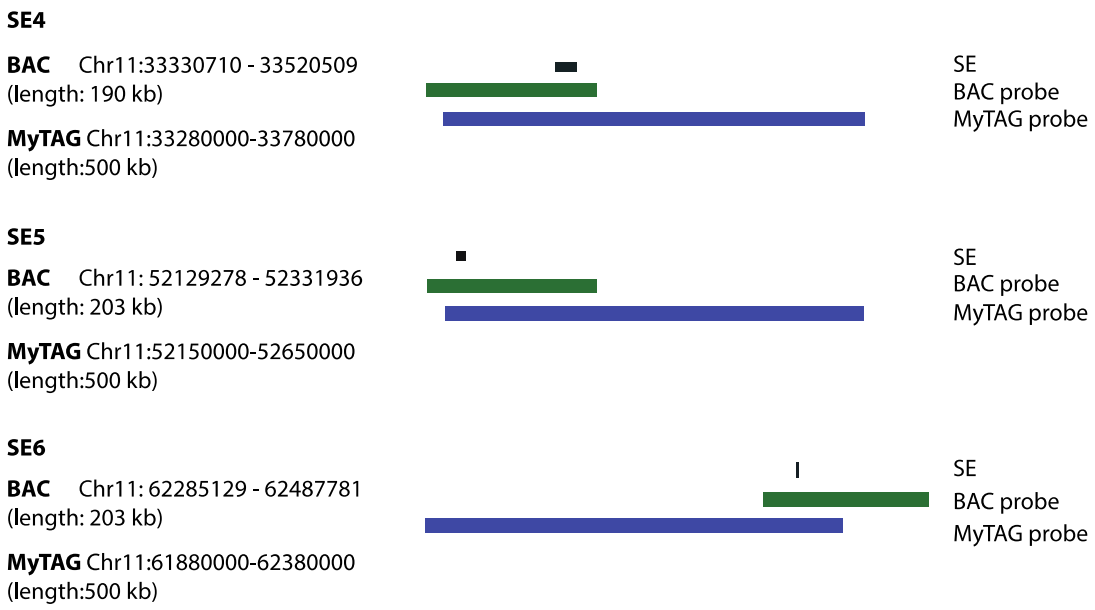


Figure 4.12: BAC probe positions in relation to the myTags probes and the super-enhancer location

Comparison of super-enhancer position, BAC probe and myTags probe. Genomic coordinates of the FISH probes are indicated on the left side (all mmg). BAC probes were chosen from a commercially available site of BAC probes, such that the overlap with the original myTags probe would be maximized.

Figure 4.13A summarizes the positions of the BAC probes on chromosomes 3, and Figure 4.13B displays example images, highlighting the interaction between BAC regions containing the super-enhancers. The enlarged images in the panels of Figure 4.13B show the contacts between the super-enhancer BAC probes. Some examples were found where FISH signals had an elongated shape, possibly extending to a part of the other super-enhancer region.

Pairwise FISH experiments on ~200 nm ESC cryosections were performed using both the BAC probes covering SE4, SE5 and SE6. I also analysed the colocalisation of S5 and SE6, a combination of probes not tested previously in Beagrie et al. 2017. Using the previously discussed pipelines for center-of-mass and edge-to-edge detection revealed a preserved high colocalization between the distal BAC probes and myTag probes (Figure 4.13C). 47 % of the BAC probes covering SE4 and SE5 were colocalized below 200 nm and 67 % of the SE5 and SE6 were within 200 nm distance. Interestingly, SE4 and SE5 were at a greater distance using the BAC probes, compared to the 500 kb myTags probes, although a high percentage of SE4 and SE5 BAC probes are in contact. These results suggest, that although the super-enhancer region mitigates in part the high contact rates, part of the close interaction between the SE-TADs may be mediated by elements outside the BAC covered region.

It remains however unclear why the smaller genomic regions covered by the SE4 and SE5 BAC probes show distances which have a much broader distribution than the original SE4/SE5 distances. It will be interesting in the future to hybridize the myTags probes together with the overlapping BAC probe to better understand how each FISH probe represents the position of genomic regions relative to each other. Interestingly, in Ferrai et al. 2010, a fosmid probe covering ~40kb was found outside of the larger signal covered by an overlapping BAC probe, suggesting that the FISH signals represent the most clustered regions of the genome covered by the probes.

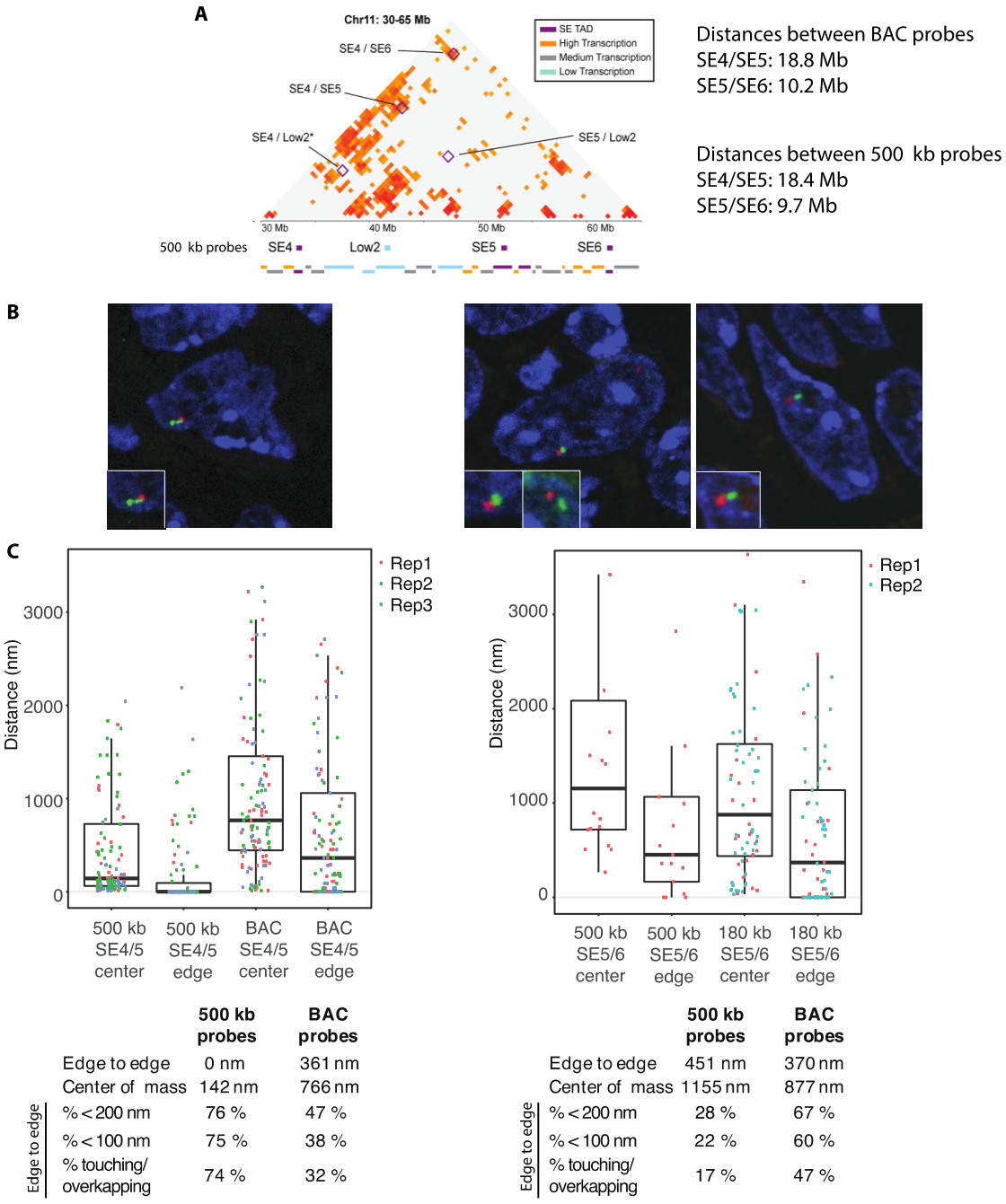


Figure 4.13: Zooming into the super-enhancer contacts on chromosome 11 shows that close proximity between super-enhancer regions is maintained

A GAM matrix showing position of myTags FISH SE4, SE5 and SE6, with respective distances on the left hand side.

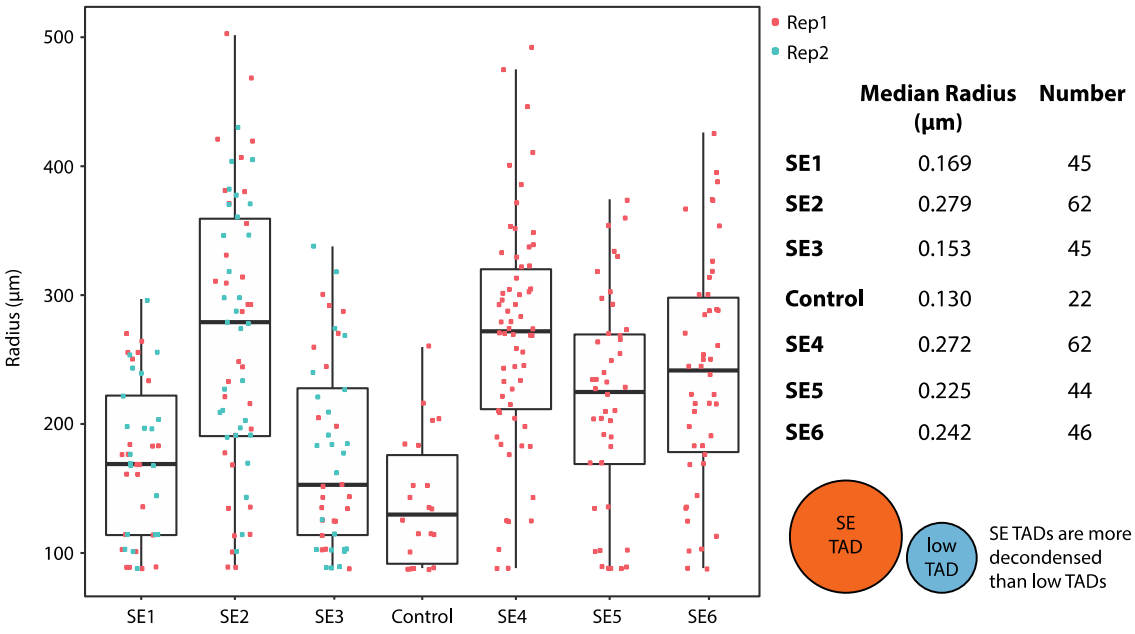
B Representative images showing contacts between SE4 (green)/SE5 (red) and SE5 (green)/SE6 (red). DAPI nuclear counterstain is shown in blue.

C Distance boxplot between the myTags probes and the BAC probes, measured both edge-to-edge and center-to-center. Results for myTag (500kb) probes for SE4 and SE4 are the same as shown in Figure 4.11 above, and replotted here for comparison. Comparing the distribution between the larger 500 kb probes and the smaller BAC probes. Three replicates were performed to test SE4/SE5 contacts between myTags and BAC probes. SE5/SE6 contacts were tested with two replicates using BAC probes and one replicate using myTags probes.

The tables contain the edge-to-edge and center-to-center distances between the 500 kb myTags probes and the smaller BAC probes, as well as the percentage of colocalization.

4.3.5 Regions covering super-enhancers are more highly decondensed then low transcribed regions

Visual inspection of cryoFISH images showed a tendency of super-enhancer covered probe regions to exhibit a higher degree of decondensation than the control probes. To quantify the degree of decondensation I measured the area occupied by the 500 kb myTags probes in the triplets on chromosome 3 and chromosome 11 and compared them to the control probe low 1. The area measurement was performed by first segmenting the images in ImageJ and afterwards using the particle analysis function in ImageJ to obtain area values. Figure 4.14 presents the radius comparison of the super-enhancer probes and control probes. All super-enhancer probes occupy a larger area, than the low probes. For example, super-enhancers SE2 and SE4 occupy more than twice the area of the low probe. There is however some variability in the degree of decondensation, as can be seen in Figure 4.14, where the three super-enhancers contained in the triplet on chromosome 11 are on average more decondensed than the super-enhancers SE1 and SE2 on chromosome 3. The span of values obtained for the median radius in SEs1-6 (cryoFISH) was between 0.153 μm and 0.279 μm, which is comparable in magnitude to the median radius of the top and bottom probes in cryoFISH: 0.269 μm for top probes and 0.165 μm for bottom probes (Figure 4.6). A higher degree of decondensation, i.e. more open chromatin is a mark of active chromatin and can be linked to super-enhancer function in the cell, as it might enable factors required for transcription enhancer to access the super-enhancer region more easily. Also it reflects the transcriptional activity within the super-enhancer TADs, which was higher compared to the low probe, as shown by mRNA-seq in Figures 5.8 and 5.9.



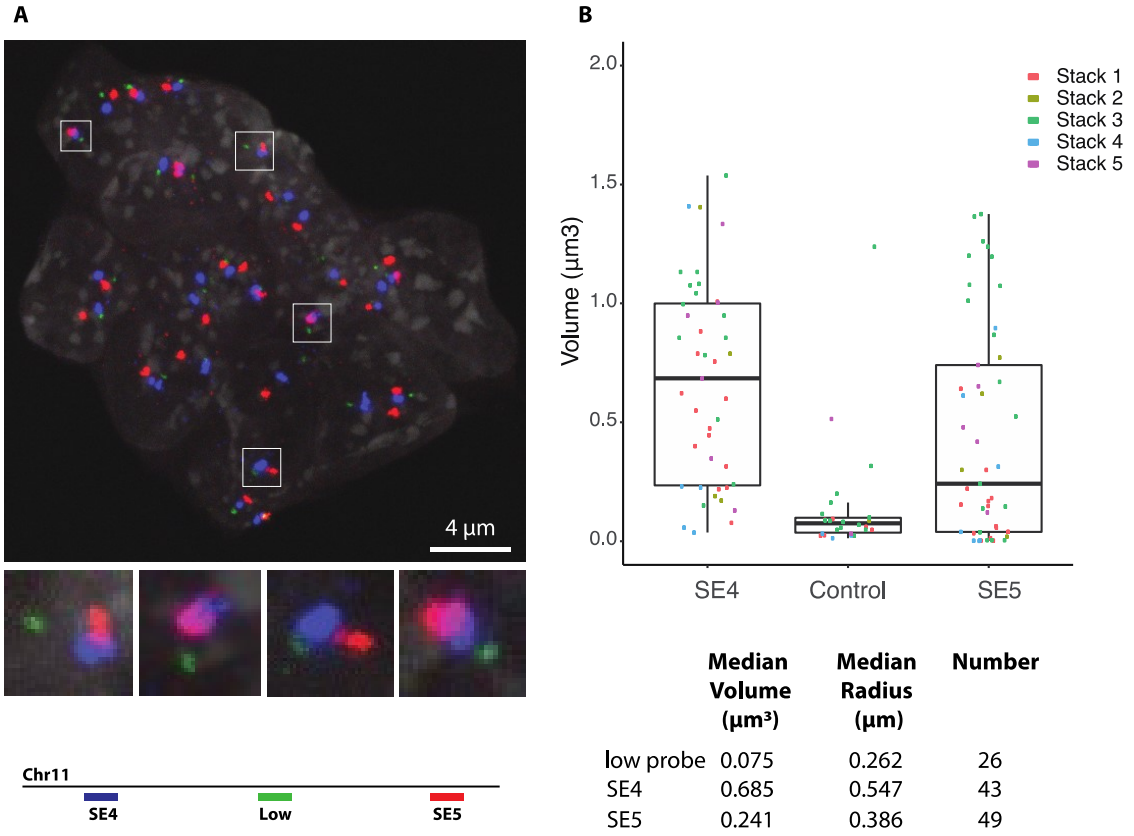
	Median Radius (µm)	Number
SE1	0.169	45
SE2	0.279	62
SE3	0.153	45
Control	0.130	22
SE4	0.272	62
SE5	0.225	44
SE6	0.242	46

Figure 4.14: Radius of super-enhancers and low control probe in cryoFISH
Radii of SEs1-6 compared to the control probe low 1. Super-enhancers are more decondensed compared to the low transcribed control probe, eg SE2 and SE4 are more than double its size.

The table on the right side contains the probe radii of the SEs 1-6 and the control probe, as well as the number of areas that were measured.

I also performed 3D FISH with the super-enhancers SE4 and SE5 that had previously been detected to be highly colocalized, and the control probe in ESC (see Figure 4.11 and Figure 4.13). Using the previously discussed Matlab pipeline (Chapter 3, Figure 3.5), I determined the probe volumes of the super-enhancers, as well as the low probe. Five 3D stacks were analysed in total to obtain the probe volumes. Figure 4.15A shows a maximum projection with SE4 (blue), SE5 (red) and the low probe (green). As can be clearly seen in the magnified images, the control probe occupied a much smaller volume than the super-enhancer TADs, despite the equal genomic length covered by all three probes. Quantitative analysis showed that the low control probe is smaller than SE4 by a factor 9.1 and smaller than SE5 by a factor 3.2. This emphasizes the strong degree of decondensation of the super-enhancer TADs in comparison to the low transcribed control region. The total number of signals detected for super-enhancers 4 and 5 is similar, however the control probe was detected less frequently, as, owing its small size, signal detection above background was not high enough to discern all FISH signals with the image analysis pipeline. Interestingly, super-enhancer 4 occupies a

larger volume than super-enhancer 5, by a factor 2.8. Also, in cryoFISH a radius difference of between super-enhancers 4 and 5 could be observed. The difference was however much smaller, in cryoFISH SE4 was larger than SE5 by a factor 1.2.



	Median Volume (µm³)	Median Radius (µm)	Number
low probe	0.075	0.262	26
SE4	0.685	0.547	43
SE5	0.241	0.386	49

Figure 4.15: 3D FISH volumes of the super-enhancers TADs compared to the control
A 3D FISH in ESC with SE4 (blue), SE5 (red) and the low control probe (green). DAPI counterstain is shown in grey. The image shows a maximum projection of a 3D stack. The enlarged images show examples of the strong decondensation of the SE probes compared to the control region. Interestingly several examples of long-range contacts between the TADs were observed.
B The plot shows the volume distribution of SE4 and SE 5 compared to the control probe. Both SE probes are strongly decondensed compared to the control, interestingly, SE4 is even more decondensed than SE5. The table contains the median volume, the median radius and the number of FISH signals measured.

Previous measurements of the pairwise distances between SE4 and SE5 in mESC cryosections showed that the super-enhancers were highly colocalized. The 3D distances between SE4, SE5 and the control probe were not yet analysed, however some examples of a strong overlap between the two super-enhancers could be detected. This suggests, that also in 3D FISH, the TADs can be observed to be frequently colocalized. Interestingly, by visual inspection, also some contacts between the low probe and SE4 were detected, which reflects the GAM data, where a non-zero contact probability for SE4/low probe 2 was detected (Figure 4.11 TableA).

4.3.6 Super-enhancer clustering in different tissues

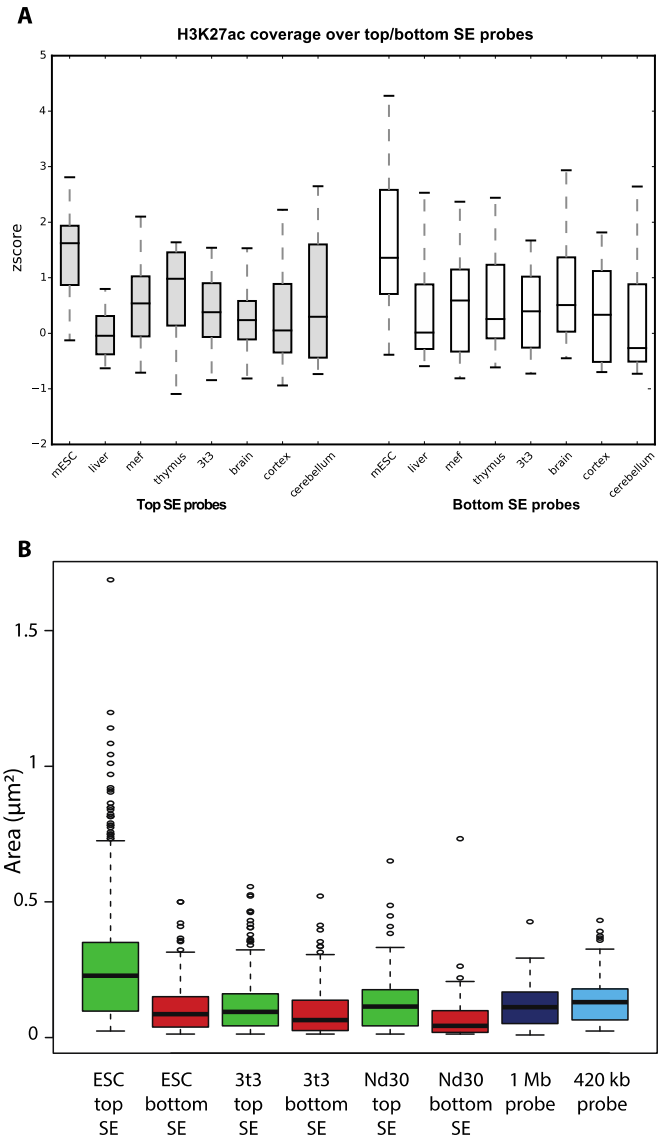
Are super-enhancer clusters stable in different cell types, or are they instead part of temporary hubs that dissociate in the process of differentiation? Using the top and bottom super-enhancer probes (Figure 4.2E), which showed extensive clustering in mESC, I performed FISH in different cell types to measure potential changes in the clustering behaviour during the differentiation. Performing one FISH experiment with top super-enhancer TAD probes (ATTO594), bottom super-enhancer TAD (ATTO550) probes and a control probe of 1 Mb length (ATTO488), on each ESC, 3T3 fibroblasts and mature, post-mitotic dopaminergic neurons on 200 nm cryosections, I tested whether clustering changes from ESC to different cell types. FISH with the top and bottom super-enhancer probes was done including a 1 Mb probe (Sox9 A, table 2.2) as a size control for clustering. To assess possible changes in super-enhancer cluster size in the different cell types, I performed area analysis as an indication of cluster size, as previously done (see Figure 4.6). Each of the super-enhancer TADs within the top and bottom probes occupies 400 kb genomic DNA and the 1 Mb probe served as a measure for the volume two and a half probes would occupy. Figure 4.15A shows H3K27ac coverage, used as a hallmark for super-enhancer activity, of the top and bottom super-enhancers in different tissues. H3K27ac coverage was measured across the super-enhancer TADs within the top and bottom super-enhancers. It is highest in mESC both in top and bottom probes and drops to lower levels in different cell types and tissues; H3K27ac analyses were performed by Robert A. Beagrie (laboratory of Ana Pombo) using published ChIP-seq data (ENCODE project in E14 mESC).

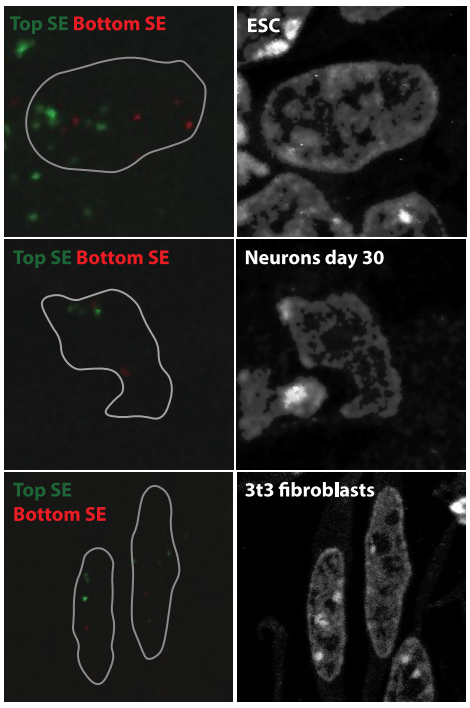
Figure 4.15C shows images of top and bottom super-enhancers green and red respectively in cryosections from ESCs, NIH-3T3 fibroblasts (labelled 3t3), and post-mitotic neurons differentiated in vitro from ESCs (labelled Nd30); The 3T3 samples were grown by Joao Dias and samples fixed and processed for cryosectioning by Sheila Xie (laboratory of Ana Pombo), neurons day 30 samples were differentiated and validated by Carmelo Ferrai and the samples were fixed and processed for cryosectioning by Sheila Q. Xie (both laboratory of Ana Pombo). In cryoFISH, the top super-enhancer probes occupied a larger area than the bottom super-enhancer probes independent of cell type (Figure 4.16, see table).
Top super-enhancer probes in ESC occupy the largest median area, 0.288 μm^2, their median area is 3.3 times larger than the area of the bottom super-enhancers in ESC with a median area of 0.086 μm^2. ESC top super-enhancer areas are furthermore roughly double the size of

the large probe 0.288 μm^2 compared to 0.112 μm^2. The median area of top super-enhancers in 3T3 fibroblasts and Neurons was smaller in comparison to the area occupied by the top super-enhancers in ESC: 0.114 μm^2 in post-mitotic Neurons day 30 and 0.095 μm^2 in 3T3 fibroblasts. In Neurons day 30, top super-enhancers were 2.65 times larger than the respective bottom super-enhancer areas (0.114 μm^2 and 0.043 μm^2). The top super-enhancer area in 3T3 fibroblasts was by a factor 1.5larger than the respective bottom super-enhancer area. Top super-enhancer area in 3T3 fibroblasts was by a factor 2.4 smaller compared to ESC (0.095 μm^2 vs 0.228 μm^2) and by a factor of 1.2 smaller compared to neurons (0.095 μm^2 vs 0.114 μm^2). In 3T3, the difference in z-score of H3K27ac between the bottom and top probes was small, which might account for the small difference in area size.

The 420 kb FISH probe, shown as a reference for a single super-enhancer expansion in Figure 4.16B, was taken from a previous experiment (table 2.2, probe A, chapter 2). FISH experiments with the single probe were done independently and not by co-hybridization with the top, bottom super-enhancer probes and the 1 Mb probe. Equally, no area comparison of the 420 kb probe across the different cell types was done. In ESC, the 420 kb probe occupies an area similar to the super-enhancer top probes in the different cell types, but also of similar dimensions as the large 1 Mb probe. As the two reference probes for clustering of ~2.5 probes and single cell probe volume have comparable median areas, it is difficult to attribute cluster sizes to the individual areas measured in different cell types.

It can be concluded, that top super-enhancers occupy the largest areas in ESC, where the median area size is roughly double the area of the 1 Mb probe and the 420 kb probe, showing some evidence for clustering in ESC. In 3T3 and post-mitotic dopaminergic neurons the median area size is reduced, however a difference in area size for bottom and top probes is maintained both in 3T3 and post-mitotic neurons, suggesting that although clustering might not be conserved, top and bottom super-enhancers maintain different states in different cell types.
The results presented here therefore suggest that super-enhancer cluster formation is dynamic and might be adjusted to the respective transcriptional needs a cell has in different cell types. This is in accordance to experiments suggesting that SE regulating cell type specific genes (Hnisz et al., 2013; Ibarra et al., 2016).





	Median area (µm²)	Number
ESC - top	0.228	899
ESC - bottom	0.086	683
Nd30 - top	0.114	184
Nd30 - bottom	0.043	99
3t3 - top	0.095	484
3t3 - bottom	0.065	314
single probe 420 kb (ESC only)	0.130	32
large 1Mb probe (ESC, Nd30, 3t3)	0.112	65

Figure 4.16: Top and bottom super-enhancer clustering changes during differentiation
A The boxplot shows H3K27ac coverage of the top/bottom SE probes in different cell types. Z-score is highest in mESC and decreases in all the other cell types. (The analyses in 5.15 A, of the z-score in different tissues was performed by R. A. Beagrie.)
B The area of the top and bottom SEs in ESC, 3T3 and Nd30, a 1Mb probe and a 420 kb probe. The 1 Mb probe was included in the FISH experiments as a reference for clustering as it contained 2.5 times the genomic content of the SE probes. The area of the single probe was taken from an independently performed experiment which was only performed in ESC.
C Some example images of the top (green) and bottom (red) SE probes in ESC, post mitotic dopaminergic neurons day 30 and 3t3 fibroblasts. The respective DAPI counterstain is shown on the right hand side.

The table contains the median areas of the top/bottom SE probes, as well as the median areas of the 1 Mb probe and the 420 kb probe. Also shown are the number of areas that were analysed per cell type.

4.3.7 Super-enhancers colocalize at splicing speckles

The nucleus is not a homogenous environment but contains nuclear landmarks, such as the nuclear lamina, the nucleolus and splicing speckles which contribute to organizing the genome. Many genomic regions are known to preferentially contact certain nuclear domains such as the lamina or the nucleolus, which contributes to the functional organization of the chromatin and aiding transcription and silencing. An interesting question regarding the clustering between super-enhancers is whether they occupy preferential positions within the nucleus, which may contribute to organising their enhancer activity. As super-enhancer regions are often highly transcribed, we hypothesized whether their clustering might coincided with their co-association with splicing speckles. For these experiments, I chose to work with thick sections; Figure 4.17A shows that the cryoFISH experiments with myTag probes work using thicker (~600 nm) cryosections.

Splicing speckles are regions of high splicing factor concentration in the nucleus (and low DNA concentration), that are involved in mRNA splicing of genes attached to their periphery (Shopland et al., 2013, Hall et al.,2006). To visualize splicing speckles in the nucleus immunofluorescence staining with the Sm-antibody that stains the spliceosomal factor Sm, can be performed. Sm is highly concentrated in speckles. To enable speckle staining on the thick 600 nm cryosections, I adapted the staining protocol and optimized the antibody concentration to achieve optimal speckle staining (Figure 4.17B). Figure 4.17B shows two example images of cryosections with 600 nm thickness. The left image is a staining with the fluorophores AlexaFluor647 and in the right image, an FITC labelled secondary antibodies was used. As can be seen from the immunofluorescence images, speckle staining coincides with small areas of depletion in the nuclear counterstain. Larger areas of nuclear counterstain depletion mark the nucleolus. Segmentation of domains with high splicing factor concentration marking the speckles was performed by thresholding the immunofluorescence signal intensity in ImageJ. The graph in Figure 4.17C shows the signal intensity along the yellow line in the image on the left. By setting a constant threshold for all images at 100, taking only into the account the signal with an intensity greater than 100, areas of high Sm concentration were segmented. The black line indicates the threshold; signal above the threshold is included in the segmented area, and signal below is discarded. The blue lines in the image 4.17C, center indicate the segmented speckle areas.

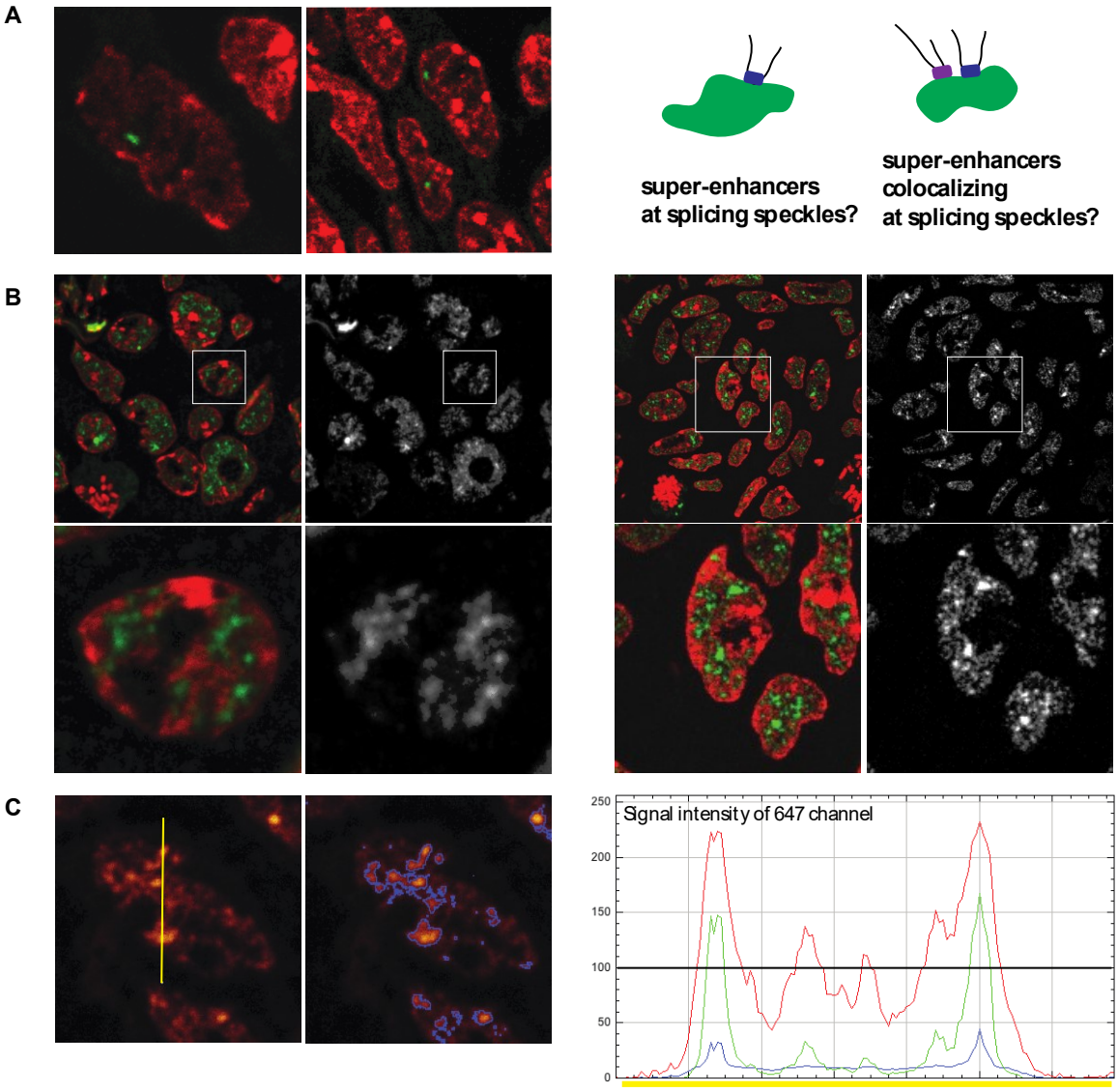


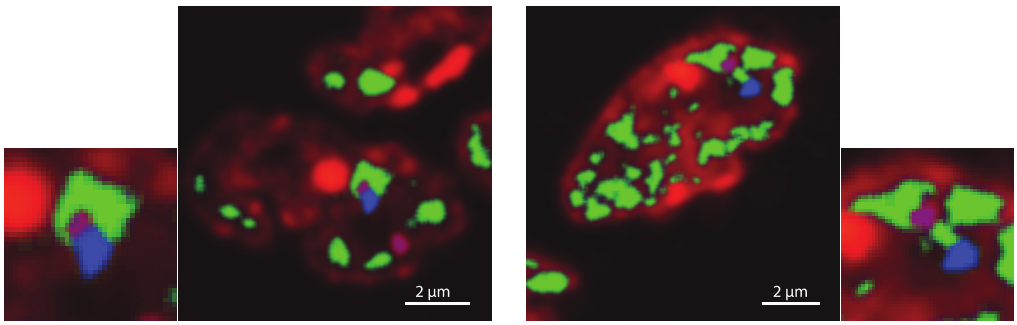
Figure 4.17: CryoFISH on 600 nm ESC cryosections shows a tendency of Super-enhancers to localize at splicing speckles

A Performing FISH with a super-enhancer probe (SE3) on 600 nm thick ESC cryosection reveals a preferred location of super-enhancers at nuclear sites occupied by splicing speckles

B Optimizing Sm speckle stainings for 600 nm thick sections with fluorophores Alexa647 (left) and FITC (right). Enlarged images show splicing factor distribution.

C Segmentation of speckle staining was based on the signal intensity. Threshold for high splicing factor enrichment was chosen to include all signal above an RGB value of 100. The image on the right shows an example for splicing factor segmentation. Signal intensity distribution along the yellow line is shown in the graph on the side.

To test localization of super-enhancers at splicing speckles, I performed one cryoFISH experiment with the super-enhancers SE5 (Alexa 488) and SE6 (Alexa 568) BAC probes (10.2 Mb apart) on 600 nm cryosection which were pre-stained with Sm-antibody with Alexa Fluor 647. Segmentation of the Sm-signal to discern speckle domains was done as described in Figure 4.17C. FISH signals were manually segmented in ImageJ. Super-enhancers were frequently found localized at speckle domains. Localization at the speckle domain occurred for 83% of super-enhancer 5 FISH signals and 70 % of super-enhancer 6 signals. In total, 43% of SE5 and SE6 were colocalised at zero distance with respect to each other, which is in line with the previously determined colocalization frequency of the SE5/SE6 BAC probes (Figure 4.13). Interestingly, all super-enhancers engaged in a contact were localized at the splicing speckle. The two images below show the 600 nm ESC cryosection with segmented splicing speckle staining and segmented FISH signals. Splicing speckles are shown in green, FISH signal of SE5 is shown in purple and SE6 in blue, the DAPI nuclear counterstain is shown in red. The image on the left shows an example of a colocalized super-enhancer contact at the speckle. The image on the right shows two SE signals which are not colocalised but nonetheless locate at the speckle. This is the first time that long-range super-enhancer contacts have been detected at splicing speckles. The occurrence of super-enhancer contacts at splicing speckles suggests that clusters of distal regulatory elements could work together with splicing speckles to provide platforms for gene regulation and splicing.



SE5 at speckle	SE5 isolated	SE6 at speckle	SE6 isolated	Colocalizaton of SE5/SE6	Colocalizaton of SE5/SE6 at speckle
83%	17%	70%	30%	43%	100%

total number of green signals: 23
total number of red signals: 27

Figure 4.18: Super-enhancers colocalize with high frequency at splicing speckles
For image analysis purposes, the splicing speckle signal and super-enhancer signals were segmented and subsequently merged. To test super-enhancer localization more directly, the shorter BAC probes were used. The super-enhancer probe SE5 is shown in purple and SE6 is shown in blue, segmented splicing speckle signal is shown in green, DAPI counterstain is red. The left image shows the two super-enhancers SE5 and SE 6 colocalizing at the splicing speckle and one super-enhancer, SE6 (purple) that is not associated with a speckle. The image on the right shows the two super-enhancers SE5 and SE6 non colocalized, but nonetheless associated with the splicing speckle.

The table shows the percentage of super-enhancers 5 and 6 either at the speckle or "isolated".
43% of the super-enhancers were coloclized and all of the colocalized super-enhancers were detected at a speckle domain.

4.4 Discussion

4.4.1 Super-enhancers form clusters of multimer contacts

Genome architecture mapping (GAM) discovered the occurrence of long-range interactions between distal regions and triplet formation between TADs containing super-enhancers. To investigate the organization of long-range interactions and the formation of multimer contacts we designed two FISH probes, each covering a set of 18 super-enhancer containing TADs. One set covered 18 super-enhancer TADs that detected a subset of the 20% most triplet forming super-enhancers, the second set contained SE TADs from the 20% bottom triplet forming probes. Performing FISH experiments both in 3D ES cells and cryosections I could show that the median area encompassed by the top super-enhancer probes is more than twice the area of the bottom probes. This allows for two possible explanations: if both top and bottom probes have the same median radius it would suggest that the top super-enhancers form median clusters of twice the number of super-enhancer containing TADs than the low TADs. An alternative interpretation takes into account different radial expansions that TADs and super-enhancer TADs can occupy. Support for this view comes from visual inspection of the 3D signal, as the 3D signal of individual bottom probe FISH foci appears smaller in radial diameter than the FISH foci of the top probes. This is also in line with their difference in transcriptional output as determined by GRO-seq, where the top super-enhancers show a higher rate of transcription than the bottom probes. Regions involved in active transcription are typically more decondensed than silent regions. By visual inspection extensive clustering was detected amongst top and bottom probes, an observation that is further supported by the high standard deviation of the top and bottom SE probe volumes. Comparing the median volume of the top and bottom super-enhancers with the radii obtained from 3D FISH SE4/SE5 on chromosome 11, shows, that the median volume of the top probes is comparable to the median volume of SE5. The bottom super-enhancers are more condensed. SE4 was the most decondensed both in 3D FISH and cryoFISH. Separating the top and bottom super-enhancer FISH volumes in multiples of their respective median volume was used as an approximation to quantify the number of super-enhancer TADs involved in cluster formation. Using this approach, I could detect extensive clustering for both top and bottom probes, with more than 20% of the top-enhancers and 30% of the bottom enhancers forming clusters of three or more interaction partners. Surprisingly, bottom probes were found to be more highly clustering, which was due to their smaller median size. On the other hand, in total 12% more

bottom super-enhancers signals were detected which argues that in general more bottom super-enhancers are found individually.

These results show that the formation of long-range contacts and clustering is inherent to both top and bottom super-enhancers. Possible differences between the top and bottom probes could be due to varying transcriptional activity induced at the respective sites, with top super-enhancers potentially triggering stronger activation, as reflected in the higher GRO-seq values. Alternatively, more highly interacting super-enhancers could interact with a greater number of different super-enhancer TADs depending on transcriptional needs. Whereas bottom super-enhancers might form more stable contacts with a smaller number of different super-enhancer containing TADs, possibly for structural purposes. It is also important to consider that the selection of the top interacting super-enhancers was chosen based on their individual interaction propensity, without accounting for the degree of mutual interactions. A possible scenario could be that the results obtained here are a snapshot of much larger cluster networks with even greater numbers of interaction partners.

As chromosomes were labelled with a maximum of three super-enhancer TADs per chromosome, clusters of more than three super-enhancers of top or bottom type imply the formation of trans-chromosomal contacts. The ~12 % of clusters with a size of 4 or more interacting super-enhancer TADs, suggests that trans contacts are frequent, amongst the set of 18 super-enhancer TADs per top and bottom probes. Trans-contacts have often been treated as background signal in Hi-C analysis, however the occurrence of chromosome intermingling has been shown in 2006 by Branco & Pombo. Branco and Pombo showed that chromosomes intermingle at their peripheries and they proposed the interchromosomal network model, which accounts for a novel level of chromosome organization by considering both intra- and inter-chromosomal interactions in maintaining chromosome stability and transcriptional activity. By correlating transcriptional inhibition with a significant decrease in chromosome intermingling, they could show that transcriptional co-regulation is a central mechanism in organizing inter-chromosomal contacts (Branco & Pombo, 2006). Trans-contacts of super-enhancers could therefore contribute to trans-chromosomal gene regulation.

An interesting approach to obtain a quantitative measure for global network sizes of super-enhancers and how the top and bottom super-enhancers connect to super-enhancer TADs at large, is a network model recently developed by Franka Rang and Tiago Rito (laboratory of

Ana Pombo). Based on Hi-C, conserved TADs (CATs) are defined between ESC and Neurons and subsequently interaction strength between those CATs is calculated by taking the mean of the Hi-C bins obtained in ESCs or Neurons that represent interactions between CATs. Using this model in a first attempt to explore differential networks of super-enhancers on chromosome 3 (SE1, SE2, SE3 and Low probe 1) showed that indeed these three super-enhancers belong to a network but do not share any interactions with the low probe and are also distant in the network, suggesting that they are located in different regions of the chromosome (Franka Rang). Extending this analysis to the super-enhancers on different chromosomes might provide new insights into how trans-chromosomal super-enhancer hubs interact and how they integrate into the neighbouring chromosome interaction networks. A similar approach was recently taken by Thibodeau et al., 2017, where using network theory models on super-enhancers, they found super-enhancers interacting more frequently with each other compared to typical enhancers (Thibodeau et al., 2017). Stating that within the networks super-enhancers exhibit hub like connectivity. Expanding on this and connecting 3D networks of super-enhancers with genome organization can generate novel insights into 3D genome architecture and the possible role enhancer hub formation has in cis and trans.

4.4.2 Super-enhancers engage in very long-distance interactions

To understand the frequency and distance between super-enhancer contacts we designed probes that cover two super-enhancer triplets with long-range interactions between the enhancer TADs. Using FISH I could show extensive colocalization between distal super-enhancers, contacts at this distance are novel and have not been previously examined. This sheds a new light on the occurrence of very long-range interactions in the magnitude of many Megabases distance and how they can contribute to gene expression regulation. One view regarding enhancer-promoter interactions, is that most occur within topological domains and TAD boundaries are often considered as a way to limit the reach of enhancers to within its TAD (Smith et al., 2016; Lupianez et al., 2015). However using GAM a large number of very long-range interactions between super-enhancer TADs and highly transcribed TADs and/or other super-enhancer TADs can be detected. With FISH I could confirm the occurrence of long-range super-enhancer contacts and show that these events occur with high frequency of more than 50% colocalization at zero distance and in close spatial proximity. All super-enhancers showed a higher percentage of colocalization in comparison to low transcribed control probes, despite the smaller distance between the super-enhancer and low probe. Using

smaller 180-200 kb BAC FISH probes, that covered the super-enhancer regions on the triplet on chromosome 11, I tested whether colocalization is maintained and hence the super-enhancers are indeed the driving element in forming the contact. The BAC probes showed an increased degree of interaction for SE5 and SE6 and a decrease of interaction for SE4 and SE5. However, the median distance between SE 4 and SE5 was within the range of SE5 and SE6. The consitently high colocalization frequency makes the super-enhancers the likely candidates for mediating the long-distance contact. Further support for super-enhancer interaction, in particular of the super-enhancers 4 and 5 came from 3D data, where visual inspection showed some examples of striking intermingling of these distal regions. The question whether triplet associations are driven by the super-enhancers themselves was also tested with GAM and it was shown that SE containing 40 kb windows co-segregate more frequently with the two other SE-TADs in their triplet than 40 kb windows located 120 kb upstream or downstream.

4.4.3 Super-enhancers are highly decondensed

One of the early observations in genome architecture was the discovery that different genomic regions are folded in varying degrees of density with the silent, heterochromatic regions more highly condensed and the active regions more open (Gilbert et al., 2004). Mechanistically this is due to the facilitated access provided by an open chromatin fiber to the transcription machinery, versus a compacted, sterically hindered genomic environment. The latter is less easily transcribed or reached by regulatory elements, whereas the former is available to binding of transcription factors and RNAPII, amongst others, thereby allowing active transcription of the genomic region to occur. Recent studies have shown evidence that different scaling factors of compaction characterize the folding of regions with different epigenetic marks, generally stating that active regions, marked by H3K4m3 show a higher degree of openness (Boettiger et al., 2016). Measuring the volume of SE FISH probes versus the control low TAD probe in cryoFISH and 3D FISH shows that super-enhancers are indeed much more decondensed, with some super-enhancers occupying almost twice the volume of the equally sized low control probe. Also amongst the super-enhancers there is some variability in the occupied volume, e.g. in 3D FISH the volume of SE4 was by a factor of 1.4 larger than SE5. Volume differences are possibly linked to super-enhancer activity and transcriptional activity within the TAD. In top and bottom probes, the lower median volume of bottom probes most likely stems from a smaller size of the individual probe regions, which

is in line with the lower transcriptional activity within those TADs. The open decondensed nature of super-enhancers fits into the picture suggested before, where they might create hubs that regulate active genes.

4.4.4 Super-enhancers clustering area decreases in different tissues, but top super-enhancers maintain a higher cluster area than the bottom super-enhancers

Some super-enhancers maintain their enhancer character in different tissues, whereas others lose thir enhancer activity. Equally, it has been reported, that novel super-enhancers can be acquired during ES cell differentiation (Bojcsuk et al., 2017). Acetylation of H3K27, a hallmark of super-enhancer occurrence, which was taken as a proxy for super-enhancer occurrence, changes in different cell types, suggesting changes in the super-enhancer state. To understand the behaviour of existing super-enhancers in different cells I evaluated their sizes as means to infer cluster formation. The results show that super-enhancer clusters for the top and bottom probes decrease in size from ESC to 3T3 fibroblasts and post-mitotic Neurons day 30. This is consistent with the decreased levels of H3K27ac observed in 3T3 fibroblasts and post-mitotic Neurons. However top probes still maintain a higher median area than the bottom probes. This shows that super-enhancers form dynamic clusters that change during differentiation, possibly as a response to the transcriptional demands of the different cell types.

As a reference for a clustered area a 1 Mb probe was added to all the experiments in ESC, post-mitotic dopaminergic neurons and 3T3. As each individual super-enhancer region of the top and bottom probes encompasses 400 kb the 1 Mb probe represent the equivalent of 2.5 clustered SE regions. In the experiment, the 1 Mb probe still occupies approximately the same area as the neurons top regions, only the ESC top probes are twice as large on average as the 1 Mb probe. This could indicate that clustering is still occurring, only it is more pronounced in ESC. On the other hand the 1Mb probe was of similar area dimensions as a probe of 420 kb length, that was used as comparison for the area of a single region. FISH with the 420 kb had been done separately in mESC cryosection, not in other cell types. Area sizes of individual super-enhancers (Figure 4.14). The results show that top and bottom super-enhancers change area size in 3T3 and neurons, compared to ESC. Also, top probes occupy larger areas, independent of cell type.

4.4.5 Super-enhancers preferentially locate at splicing speckles

Splicing speckles are nuclear landmarks enriched in splicing factors, they are typically located between the chromosome territories (Hall et al., 2006). It has been reported, that multiple genes, which can also originate from different chromosomes, can locate simultaneously at speckles and mRNA of the respective genes has been found within the speckle interior (Shopland et al., 2003). FISH of two highly interacting super-enhancers on 600 nm cryosections revealed a preferential localization of super-enhancers at speckles and a high incidence of super-enhancer contacts at the speckle domain. These findings provide support for a model, where speckles are at the center of hubs of transcriptional activity, that potentially cluster active genes and their regulatory regions, with RNAPII and the splicing machinery. As FISH with top and bottom probes suggested the occurrence of trans-chromosomal contacts, it is tempting to speculate that these also coincide with speckle domains. This would also fit with the interchromosomal network model by Branco and Pombo (2008), which proposed the presence of nuclear landmarks that provide tethering to intermingling regions from neighbouring chromosomes. Possibly speckles are central in organizing transcription by creating hubs that super-enhancers, active genes and transcription factories can associate with (see Figure 4.18).

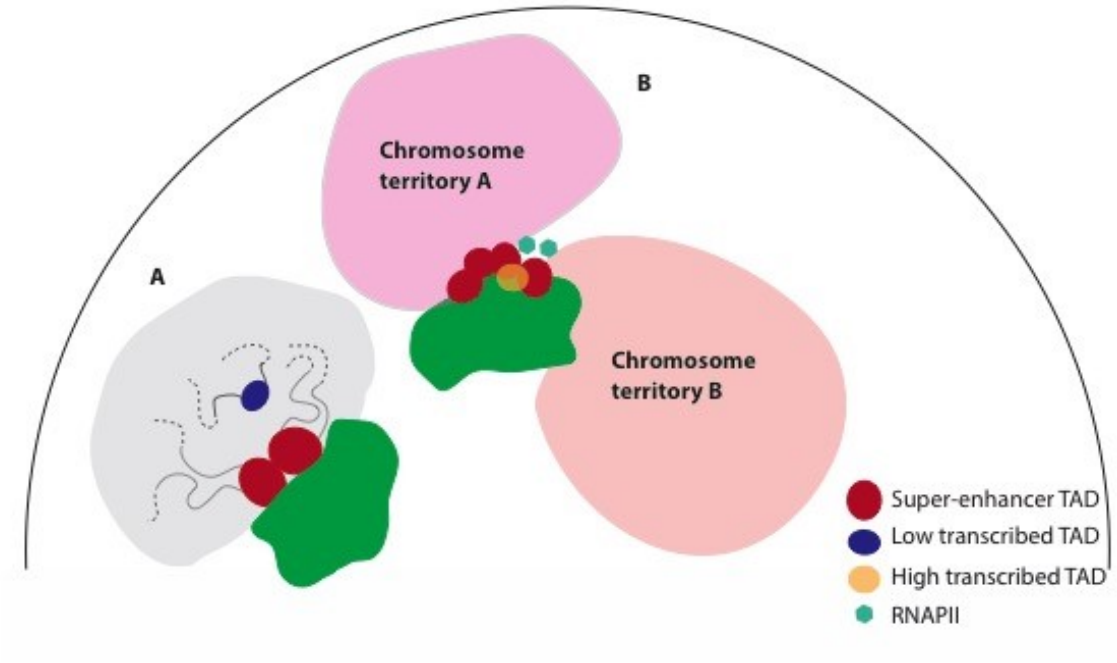


Figure 4.19: Potential organization of long-distance super-enhancer interactions
A Data showed that super-enhancers are located at a closer median distance and are more often colocalized than with the control probe. Furthermore super-enhancers are found colocalized at the splicing speckle. **B** Possibly clusters of super-enhancers from neighbouring chromosomes locate at the splicing speckle, where they could provide regulatory platforms for active genes. These hubs could colocalize with transcription factories, creating domains of high transcriptional activity and long-range gene regulation.

4.4.6 Conclusion and outlook

This work showed that super-enhancers engage in very-long range interactions with high frequency. The frequency of interaction does not always directly relate to the distance between the interacting TADs, with some far-away regions exhibiting higher colocalization frequencies than closer regions. This demonstrates that the genome is organised into functional domains, also at great large distances. Using smaller BAC probes to investigate which parts of the TAD are engaged in the distal contacts, showed that the super-enhancer regions are crucial in mitigating the interaction. Furthermore, evidence for the formation of triplets and clusters of more than three interaction partners has been shown. Cluster size and clustering between top and bottom super-enhancer probes also suggests that super-enhancers are involved in the formation of trans-chromosomal contacts. SE clusters are dynamic and change during the differentiation. Compared to low transcribed probes super-enhancers are in highly decondensed regions and finally I present indications of their preferential location at

splicing speckles possibly to mechanistically co-associate with active genes, to take advantage of subsequent mRNA processing by the speckles.

Further experiments with FISH probes, covering super-enhancers that have high GAM Pi interaction values on sections, pre-stained with splicing speckles should be performed to test, whether colocalization at the speckle domain is a frequent phenomenon amongst super-enhancers. To expand on the idea of transcriptional hub formation it will be important to test whether simultaneous colocalization of super-enhancers and active genes occurs. Here perturbation experiments where transcription is blocked could also provide insights into understanding whether clustering between super-enhancers is maintained. Simultaneous immunofluorescence of transcription factories and RNA FISH could provide new insights into whether active transcription occurs at these sites. To improve detection of splicing speckles and transcription factories it would be interesting to apply super-resolution methods, such as STED, which has previously been tested on cryosections and shown significant improvements in resolution (Appendix, Figure 7.3). It would furthermore be interesting to investigate the formation of trans-chromosomal contacts, by ranking super-enhancers that have high probability of interaction values for trans-chromosomal interactions and test their colocalization with FISH. Also utilizing network models to investigate the role of trans-chromosomal super-enhancer hubs and their possible roles in transcription regulation and structural tethering of neighbouring chromosomes will be fascinating to use for further investigations.

5 Summary and Discussion

5.1 Summary

In this work, I investigated different levels of chromatin architecture to understand how they contribute to the 3D folding of chromatin and how this can influence nuclear function. On the smallest scale, I investigated TAD heterogeneity and the organization of two neighbouring TADs, one containing the Sox9 locus (Sox9 TAD) and its adjacent TAD containing the Kcnj genes (Kcnj TAD). TAD organization and reorganization are crucial in the context of genetic disease, as reorganization can give rise to novel contacts that may initiate ectopic gene expression.

Many, naturally occurring duplications have been known to occur around the Sox9 locus, some of which are associated with disease (Lybaek et al., 2013; Franke et al., 2016). To further understand how duplications can alter the original locus structure, I performed FISH on cells that have a duplication spanning the Sox9-Kcnj TAD boundary. Here I showed how local genome architecture reorganizes as a result of the genetic alteration and two populations arise, with different Sox9 Duplication arrangements.

I also tested the occurrence of very-long range interactions between distal TADs containing super-enhancers. I found evidence for cluster formation amongst super-enhancers and confirmed frequent interactions between super-enhancer TADs, as seen in GAM data. Volume measurements of the super-enhancers regions showed a strongly decondensed nature, which is in line with their high activity and suggests a possible facilitated access of genes that interact with these enhancer elements. I also showed that cluster formation is decreased in other cell types, however super-enhancers that were characterized as highly interacting in ESCs maintained a higher median (cluster) area than their lowly interacting counterparts. I was also able to detect frequent localization of super-enhancers at splicing speckles, as well as frequent colocalization of two distal super-enhancers at splicing speckles.

5.2 Methods comparison – cryoFISH, 3D FISH, Hi-C and GAM

Different techniques are available to investigate 3D genome architecture and in this work data from cryoFISH, 3D FISH, GAM, Hi-C and cHi-C was integrated to investigate local chromatin architecture and distal inter-TAD contacts. Comparison of the results obtained by different methods shows that many similar features are detected, for example Chromosome capture methods, as well as GAM detect the formation of TADs. FISH measurements of distances between genomic regions are larger when they are separated by a TAD boundary than two regions within the same TAD (Nora et al., 2012; Dixon et al., 2012; Beagrie et al., 2017). At the same time, methods differ substantially in their detection of long-range contacts and their ability to capture multimer contacts and single-cell variability. GAM data has shown that long-range contacts occur with high frequency, at previously unknown length scales, a finding that was also observed by FISH but not Hi-C. A possible reason for the difference in detection of long-range contacts between GAM, FISH and Hi-C may be the difference in chemical fixation. CryoFISH and GAM are based on a strong fixation in 4% depolymerized PFA followed by 8% for 1-2 h, which is highly stringent and was shown to lead to optimal structural preservation (Pombo et al., 1999, Guillot et al., 2004, Pombo, 2007). In Hi-C, on the other hand, fixation is much weaker, typically 1-2 % formaldehyde, or even completely omitted (Rao et al., 2017; Rao et al., 2014). Low fixation has however previously been shown to introduce artefacts (Guillot et al., 2004). Fixation preserves the cellular architecture, as well as proteins and other macromolecules in their spatial relationship to the nucleus, so positional information is maintained. Therefore, fixation is of great consequence in preparing samples for imaging and any artefacts and disruptions that are introduced by the fixative will influence downstream results. Crosslinking fixatives, such as paraformaldehyde, act by introducing covalent bonds between proteins. Formaldehyde reacts with primary amines and amides, transforms them into reactive groups which then react with other amino acids to form covalent bonds (Mason & O'Leary, 1991; Thavarajah et al., 2012). Fixation therefore imparts a (dense) mesh of covalent bonds onto the cellular ultrastructure, that acts as a scaffold for the cellular and nuclear structure. The fixation used by cryoFISH has been shown to preserve ultrastructure optimally (Pombo 2007) using weaker fixatives has on the other hand been demonstrated to disrupt the nuclear architecture and induce the formation of artefacts (Guillot et al., 2004). Especially more transient, fragile interactions may not be stable enough to be captured after a weak fixation, which might explain the lack of long-range interactions in Hi-C compared to GAM. Other experimental issues such as

restriction enzyme efficiency in Hi-C may further impact interaction frequencies and contribute to possible differences between the different methods. This shows that although the use of Hi-C greatly aids in understanding chromatin architecture it may underestimate the formation of long-range contacts and measured frequencies are not directly translatable into spatial distances.

Recent efforts to reconcile FISH and Hi-C showed that, although both methods give in many cases comparable results, there are still many discrepancies (Williamson et al., 2014; Fudenberg & Imakaev, 2017). In Williamson et al. (2014), it was for example shown that during gene activation in ESC differentiation elevated 5C cross-linking frequencies are detected, but within the same region FISH shows decompaction. This suggests that interaction frequencies do not simply reflect physical distances and Hi-C interaction frequencies cannot be used interchangeably with median FISH distances (Williamson et al., 2014). Furthermore, FISH also captures single-cell variability, which population-based methods, such as Hi-C fail to detect. For example, FISH inter-probe distance analysis can quantify the number of loci within a population that contacts regions across TAD boundaries, whereas population-based methods may lose this information due to averaging interaction frequencies over millions of cells.

In this work, two FISH methods were used, cryoFISH and 3D FISH. 3D FISH relies on a milder fixation than cryoFISH and harsh treatments that, as for Hi-C, may disturb chromatin contacts and the structures of chromosomes. Consistently, it was found that inter-probe distances in 3D FISH are longer than the equivalent distances in cryoFISH. In parallel, the FISH signal volumes were more decondensed. When comparing cryoFISH and 3D FISH distances, it is therefore important to consider potential differences that are intrinsic to the respective methods. These differences may originate from different sources, of which one important factor is likely to be again the different chemical treatment of the cells in the process of cellular fixation. With the less stringent fixation in 3D FISH, cells are more likely to undergo structural changes during the denaturation step in the FISH experiment. CryoFISH furthermore has the advantage of being a high-resolution method, with an increased z-resolution that is limited to the section thickness. In comparison to lower resolution 3D images, cryoFISH is ideally suited to determine genomic distances.

5.3 TADs are heterogeneous and TAD boundaries are permissive

TADs are heterogeneous structures with great variability of folding on a single cell level. The majority of interactions occur within the TAD, but over 10% of interactions across the TAD boundary reach the distal part of the neighbouring TAD, highlighting the permissible nature of TAD boundaries. There is much ongoing debate about the nature and significance of TAD boundaries. Several biological parameters are correlated with the TAD boundaries, such as early/late replication and the borders might be crucial in maintaining stable replication programs (Pope et al., 2014). TADs and subTAD boundaries can also be flexible in the extent with which they block or permit cell type-specific chromatin interactions and gene expression (Narenda et al., 2016). TADs with lower TAD boundary strength possibly allow more interTAD interactions (Narenda et al., 2016). In the light of previous studies, my work suggests that TAD boundaries represent a population average. A majority of cells show a preference for intraTAD interactions, but a significant number of cells engage in preferential interTAD interactions. The occurrence of interactions with neighbouring TADs underscores the importance of the locus organization, and alterations within the direct genomic neighbourhood can potentially have detrimental effects on the TAD structure and respective gene expression.

5.4 Genomic TAD boundary duplication can form a mixture of differently organized TADs

TAD boundary elements are important in organizing chromatin interactions and recent studies have shown that deletions that include TAD boundaries can result in ectopic interactions that induce gene expression (Lupianez et al., 2015). Further evidence was presented by Nora and colleagues who found that a 58 kb deletion in the Xist region which removed a TAD boundary also lead to partial merging of the adjacent TADs, causing ectopic interactions and misregulation of the genes within the TADs (Nora et al., 2012). These effects have also been proven *in silico*, where polymer modelling confirmed that disrupting intraTAD interactions can cause an increase in interTAD interactions (Giorgetti et al., 2014). Franke et al. (2016) reported that the introduction of a TAD boundary at the Sox9 locus causes the formation of a separated neoTAD in developing limb buds, where the neoTAD includes the duplicated region. Using 3D FISH and cryoFISH techniques on this locus, I showed that multiple configurations co-occur in an mESC line that contained the duplication.

A majority of Sox9 loci adopted a configuration where the duplicated TAD containing the Kcnj TAD and Sox9 TAD, merges with the Sox9 TAD, forming one large TAD region. A subpopulation of loci showed the neoTAD organization, where the duplicated region forms an independent TAD. This result also highlights the single cell variability in conformations that genomic regions can acquire. Average configurations are ensembles of structures, instead of fixed mono-states. This may introduce an additional layer of complexity in the analysis of genomic duplications in the context of genomic disease, as in an organism affected by a genomic duplication, multiple configurations can be acquired that might each have different effects on gene expression. It is therefore crucial to develop 3D mapping approaches that detect and can quantify single cell behaviours in chromatin folding.

5.5 Genomic interactions reach far beyond the TAD

Gene regulation and genome organization is not limited to the TAD level. TADs engage in many interTAD interactions, ranging from interactions between neighbouring TADs to long-range interTAD interactions between regions that are many Mb apart. Most strikingly, using FISH, I could detect long-range interactions between distal TADs containing super-enhancer elements and went on to show that super-enhancer TADs colocalize with high frequency in close spatial proximity. Using smaller BAC probes that were centred on the enhancer element, I could confirm the occurrence of long-range super-enhancer interactions.

Contacts between regulatory regions that are many Mb apart are novel and have not previously been reported. Long-rang interactions of this length scale shine a new light on long-range gene regulation within the nucleus. Additionally, triplet formation between super-enhancers could be detected as well as extensive cluster formation of potentially six or more enhancers. This suggests that super-enhancers form hubs and interactions can occur in cis and measurements suggest, also in trans, i.e. super enhancers from neighbouring chromosomes may engage in a common hub. Hub formation could create the possibility of spatially concentrating transcription and the associated transcription mediating protein factors.

The localization of genomic regions from different chromosomes at the same splicing speckle has been previously reported (Shopland et al., 2003). To further this, I used FISH on two interacting super-enhancers, on mESC cryosections pre-stained with anti-Sm antibodies to

visualize splicing speckles, and showed frequent localization of the individual as well as the colocalized super-enhancers at splicing speckles. This is proposing a model where hubs of super-enhancers cluster in cis and trans at splicing speckles.

The decondensed super-enhancers might serve as platforms for active genes to locate at and become activated. Splicing speckles can provide a reservoir of splicing factors and possibly transcripts are transferred into speckles for further processing. Future FISH experiments testing this hypothesis and combining super-enhancer FISH with RNAPII staining for transcription factories might provide insights into novel chromatin organization levels. Enhancer clustering in hubs can also explain cooperative features that have been previously reported, such as: A single enhancer is able to activate simultaneously multiple genes and enhancers typically activate genes in bursts, with two genes, that are simultaneously activated by the same enhancer, having synchronized transcriptional bursts (Fukaya et al., 2016).

5.6 Compaction of genomic regions is highly variable, with more active regions showing a tendency stronger decondensation

In this work I have shown that different genomic regions can occupy very different volumes which can be an indication of transcriptional activity (Davie et al., 2015). Regions that are more open and decondensed are accessible for binding of transcription factors and RNAPII, in turn they tend to be more transcriptionally active. Even within the direct genomic neighbourhood, differences in the degree of compaction can occur, e.g. at the Sox9 locus, in mESC, a 1Mb probe spanning the gene desert, containing many regulatory elements, was much more compacted than a genomic region within the same TAD covering a 420 kb region including the Sox9 gene. Differences in compaction within a TAD also highlight the different genetic environments that co-occur within a TAD. A possible explanation could be that the 1 Mb probe covers a number of genetic elements that are not active at the ESC stage and tight compaction might prevent the enhancers from ectopic gene expression at the ES cell stage.

TADs containing super-enhancers showed a high degree of decondensation. Enhancers and super-enhancer engage in looping interactions and in particular super-enhancers are marked by a high degree of transcription factor binding, requiring a high degree of chromatin accessibility for protein binding**.** However, super-enhancers show varying degrees of

decondensation and comparing the median volumes of 18 highly interacting super-enhancers and 18 less highly interacting super-enhancers in mESC showed that the latter are more compact on average. Equally, the volumes of individual super-enhancers showed a range of different volumes. However, all super-enhancers were significantly more decondensed then the lowly transcribed probes, emphasizing the significance of compaction in relation to the activity of the TAD.

Besides global differences in transcription due to compaction which may contribute to the overall transcription state of a domain, differential local chromatin density at gene loci can also be significant. Recent studies propose that differences in local chromatin density lead to differential gene expression (Giorgetti et al., 2014; Golkaram et al., 2017) and could be at the root of transcriptional bursting and single cell heterogeneity in ESC (Golkaram et al., 2017). Possibly due to variability in distances between regulatory sequences. This emphasizes the importance single-cell heterogeneity has on the cellular state of transcription state.

5.7 Formation of super-enhancer domains by phase separation

I have discussed the possible advantages of super-enhancer clustering in providing hubs that active genes can locate to and undergo transcription and splicing. One important question is, how can these regions find each other in 3D space, especially considering the large distances in between some interacting super-enhancers, or when undergoing trans-interactions across chromosomes. Recently the formation of heterochromatin domains was proposed to occur by phase separation. Like super-enhancer clusters, heterochromatin domains are membrane less domains where genomic regions are spatially compartmentalized and segregated from their surroundings. Heterochromatin domain formation in *Drosophila* was recently described by Gros et al., 2017, explaining that these domains exhibit dynamics that are characteristic of liquid phase-separation, including reduced diffusion, increased coordinated movement and inert probe exclusion at the domain boundary. Phase separation, is the conversion of a single-phase system into a multiphase system, for example the separation of a solution into two immiscible liquids.

One of the fundamental principles of organizing nuclear activities is by means of spatial clustering of regions that cooperatively achieve a shared state or function. Be it clustering of silent heterochromatic domains, that is kept transcriptionally silent, or organizing

transcription at transcription factories, the local agglomeration of activities allows greater efficiency and stability of the involved process. How this spatial clustering is achieved mechanistically, is however non-trivial, as free diffusion within the nucleus would favour a homogenous distribution of all molecules. It is known, that membrane less domains, nuclear bodies, such as splicing speckles and Cajal bodies, are formed by phase separation, based on the cooperative interaction of multivariant molecules (Banani et al., 2017; Ban- jade et al., 2015; Bergeron-Sandoval et al., 2016; Brangwynne et al., 2009). A high density of proteins and nucleic acids, and cooperative interactions among the different molecules, leads to the formation of larger domains. Recently, the physical concept of phase separation has been successfully applied to the formation of heterochromatin domains (Strom et al., 2017; Larson et al., 2017). In heterochromatin, compaction alone does not account for the formation of the domains and fast diffusion of proteins within the domain, amongst other features (Strom et al., 2017). Instead, liquid phase separation successfully describes domain dynamics, reduced diffusion, domain boundary formation and other characteristics.

Phase separation was recently suggested as a model by which individual super-enhancers, consisting of several enhancer elements are formed and how they cooperatively achieve high transcriptional bursting, as well as simultaneous activation of several genes (Hinsz et al., 2017). Super-enhancers are occupied by high densities of transcription factors, RNAPII, chromatin regulators and cofactors, all of which adds up to an estimated ten-fold increased density of the factors compared to normal sets of enhancers (Hinsz et al., 2017). Many of these factors that bind to super-enhancers in high concentration can undergo chemical modification, such as phosphorylation, making reversible binding interactions with histones and other proteins possible. Temporary cross-linking between the molecules was suggested to induce phase separation in super-enhancers as a unit (Hinsz et al., 2017). The requirements that induce phase separation are possibly also fulfilled for hubs of super-enhancers. High densities of protein factors that are surrounding the cluster of super-enhancers could interact and build a mesh that maintains the hub integrity.

An additional fact that may positively contribute to a phase separation mechanism of super-enhancer cluster formation is the effect of molecular crowding, which has the effect of increasing effective local concentrations. Molecular crowding occurs in situations of very high macromolecular concentrations, whereby the volume of solvent available for other molecules is reduced. The effect changes the diffusion kinetics of the molecules as well as

the equilibrium constants and dissociation constants. Dissociation constants in a molecular crowding situation favour the association of molecules, in particular the formation of macromolecules, such as protein complexes, or DNA-binding of proteins to their cognate genomic sequences (Zhou et al., 2008). Molecular crowding effects and stable hub formation, supported by a mesh of inter-protein interactions could possibly also explain the recently published observation that after cohesin removal hundreds of colocalizations between super enhancers within and across chromosomes were observed (Rao et al., 2017). Alternatively, localization of hubs of super-enhancers at nuclear landmarks, such as splicing speckles could allow the super-enhancers to take advantage of the exclusive environment created by the speckles.

6 References

Allen, B. L. and D. J. Taatjes (2015). "The Mediator complex: a central integrator of transcription." Nat Rev Mol Cell Biol **16**(3): 155-166.

Andrey, G. and S. Mundlos (2017). "The three-dimensional genome: regulating gene expression during pluripotency and development." Development **144**(20): 3646-3658.

Banani, S. F., H. O. Lee, A. A. Hyman and M. K. Rosen (2017). "Biomolecular condensates: organizers of cellular biochemistry." Nat Rev Mol Cell Biol **18**(5): 285-298.

Banerji, J., S. Rusconi and W. Schaffner (1981). "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences." Cell **27**(2 Pt 1): 299-308.

Banjade, S., Q. Wu, A. Mittal, W. B. Peeples, R. V. Pappu and M. K. Rosen (2015). "Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck." Proc Natl Acad Sci U S A **112**(47): E6426-6435.

Barbieri, M., J. Fraser, L. M. Lavitas, M. Chotalia, J. Dostie, A. Pombo and M. Nicodemi (2013). "A polymer model explains the complexity of large-scale chromatin folding." Nucleus **4**(4): 267-273.

Barbieri, M., S. Q. Xie, E. Torlai Triglia, A. M. Chiariello, S. Bianco, I. de Santiago, M. R., D. Rueda, M. Nicodemi and A. Pombo (2017). "Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells." Nat Struct Mol Biol **24**(6): 515-524.

Beagrie, R. A. and A. Pombo (2016). "Gene activation by metazoan enhancers: Diverse mechanisms stimulate distinct steps of transcription." Bioessays **38**(9): 881-893.

Beagrie, R. A., A. Scialdone, M. Schueler, D. C. Kraemer, M. Chotalia, S. Q. Xie, M. Barbieri, I. de Santiago, L. M. Lavitas, M. R. Branco, J. Fraser, J. Dostie, L. Game, N. Dillon, P. A. Edwards, M. Nicodemi and A. Pombo (2017). "Complex multi-enhancer contacts captured by genome architecture mapping." Nature **543**(7646): 519-524.

Beagrie,R. A. PhD tesis (2015). "Deciphering mechanisms of gene regulation through strategies for mapping genome architecture". https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiewcWWxJPaAhWKKVAKHYXCB5sQFggpMAA&url=https%3A%2F%2Fspiral.imperial.ac.uk%2Fbitstream%2F10044%2F1%2F55112%2F1%2FBeagrie-R-2015-PhD-Thesis.pdf&usg=AOvVaw0yfg7ppjLaSzj9vSJeI2to

Bergeron-Sandoval, L. P., N. Safaee and S. W. Michnick (2016). "Mechanisms and Consequences of Macromolecular Phase Separation." Cell **165**(5): 1067-1079.

Berlivet, S., D. Paquette, A. Dumouchel, D. Langlais, J. Dostie and M. Kmita (2013). "Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs." PLoS Genet **9**(12): e1004018.

Bickmore, W. A. and B. van Steensel (2013). "Genome architecture: domain organization of interphase chromosomes." Cell **152**(6): 1270-1284.

Boettiger, A. N., B. Bintu, J. R. Moffitt, S. Wang, B. J. Beliveau, G. Fudenberg, M. Imakaev, L. A. Mirny, C. T. Wu and X. Zhuang (2016). "Super-resolution imaging reveals distinct chromatin folding for different epigenetic states." Nature **529**(7586): 418-422.

Bojcsuk, D., G. Nagy and B. L. Balint (2017). "Inducible super-enhancers are organized based on canonical signal-specific transcription factor binding elements." Nucleic Acids Res **45**(7): 3693-3706.

Branco, M. R., T. Branco, F. Ramirez and A. Pombo (2008). "Changes in chromosome organization during PHA-activation of resting human lymphocytes measured by cryo-FISH." Chromosome Res **16**(3): 413-426.

Branco, M. R. and A. Pombo (2006). "Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations." PLoS Biol **4**(5): e138.

Branco, M. R. and A. Pombo (2007). "Chromosome organization: new facts, new models." Trends Cell Biol **17**(3): 127-134.

Brangwynne, C. P., C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Julicher and A. A. Hyman (2009). "Germline P granules are liquid droplets that localize by controlled dissolution/condensation." Science **324**(5935): 1729-1732.

Brookes, E. and A. Pombo (2009). "Modifications of RNA polymerase II are pivotal in regulating gene expression states." EMBO Rep **10**(11): 1213-1219.

Brookes, E., I. de Santiago, D. Hebenstreit, K. J. Morris, T. Carroll, S. Q. Xie, J. K. Stock, M. Heidemann, D. Eick, N. Nozaki, H. Kimura, J. Ragoussis, S. A. Teichmann and A. Pombo (2012). "Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs." Cell Stem Cell **10**(2): 157-170.

Brunetti-Pierri, N., J. S. Berg, F. Scaglia, J. Belmont, C. A. Bacino, T. Sahoo, S. R. Lalani, B. Graham, B. Lee, M. Shinawi, J. Shen, S. H. Kang, A. Pursley, T. Lotze, G. Kennedy, S. Lansky-Shafer, C. Weaver, E. R. Roeder, T. A. Grebe, G. L. Arnold, T. Hutchison, T. Reimschisel, S. Amato, M. T. Geragthy, J. W. Innis, E. Obersztyn, B. Nowakowska, S. S. Rosengren, P. I. Bader, D. K. Grange, S. Naqvi, A. D. Garnica, S. M. Bernes, C. T. Fong, A. Summers, W. D. Walters, J. R. Lupski, P. Stankiewicz, S. W. Cheung and A. Patel (2008). "Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities." Nat Genet **40**(12): 1466-1471.

Campbell, P. J., P. J. Stephens, E. D. Pleasance, S. O'Meara, H. Li, T. Santarius, L. A. Stebbings, C. Leroy, S. Edkins, C. Hardy, J. W. Teague, A. Menzies, I. Goodhead, D. J. Turner, C. M. Clee, M. A. Quail, A. Cox, C. Brown, R. Durbin, M. E. Hurles, P. A. Edwards,

G. R. Bignell, M. R. Stratton and P. A. Futreal (2008). "Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing." Nat Genet **40**(6): 722-729.

Chambeyron, S., N. R. Da Silva, K. A. Lawson and W. A. Bickmore (2005). "Nuclear re-organisation of the Hoxb complex during mouse embryonic development." Development **132**(9): 2215-2223.

Cisse, II, I. Izeddin, S. Z. Causse, L. Boudarene, A. Senecal, L. Muresan, C. Dugast-Darzacq, B. Hajj, M. Dahan and X. Darzacq (2013). "Real-time dynamics of RNA polymerase II clustering in live human cells." Science **341**(6146): 664-667.

Cmarko, D., P. J. Verschure, T. E. Martin, M. E. Dahmus, S. Krause, X. D. Fu, R. van Driel and S. Fakan (1999). "Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection." Mol Biol Cell **10**(1): 211-223.

Conaway, R. C. and J. W. Conaway (2011). "Function and regulation of the Mediator complex." Curr Opin Genet Dev **21**(2): 225-230.

Consortium, E. P. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Corden, J. L., D. L. Cadena, J. M. Ahearn, Jr. and M. E. Dahmus (1985). "A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II." Proc Natl Acad Sci U S A **82**(23): 7934-7938.

Craddock, N. and M. J. Owen (2010). "Molecular genetics and the relationship between epilepsy and psychosis." Br J Psychiatry **197**(1): 75-76; author reply 76.

Cremer, T. and C. Cremer (2001). "Chromosome territories, nuclear architecture and gene regulation in mammalian cells." Nat Rev Genet **2**(4): 292-301.

Croft, J. A., J. M. Bridger, S. Boyle, P. Perry, P. Teague and W. A. Bickmore (1999). "Differences in the localization and morphology of chromosomes in the human nucleus." J Cell Biol **145**(6): 1119-1131.

Davie, K., J. Jacobs, M. Atkins, D. Potier, V. Christiaens, G. Halder and S. Aerts (2015). "Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling." PLoS Genet **11**(2): e1004994.

Di Micco, R., B. Fontanals-Cirera, V. Low, P. Ntziachristos, S. K. Yuen, C. D. Lovell, I. Dolgalev, Y. Yonekubo, G. Zhang, E. Rusinova, G. Gerona-Navarro, M. Canamero, M. Ohlmeyer, I. Aifantis, M. M. Zhou, A. Tsirigos and E. Hernando (2014). "Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes." Cell Rep **9**(1): 234-247.

Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu and B. Ren (2012). "Topological domains in mammalian genomes identified by analysis of chromatin interactions." Nature **485**(7398): 376-380.

Dowen, J. M., Z. P. Fan, D. Hnisz, G. Ren, B. J. Abraham, L. N. Zhang, A. S. Weintraub, J. Schujiers, T. I. Lee, K. Zhao and R. A. Young (2014). "Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes." Cell **159**(2): 374-387.

Ferrai, C., E. Torlai Triglia, J. R. Risner-Janiczek, T. Rito, O. J. Rackham, I. de Santiago, A. Kukalev, M. Nicodemi, A. Akalin, M. Li, M. A. Ungless and A. Pombo (2017). "RNA polymerase II primes Polycomb-repressed developmental genes throughout terminal neuronal differentiation." Mol Syst Biol **13**(10): 946.

Ferrai, C., S. Q. Xie, P. Luraghi, D. Munari, F. Ramirez, M. R. Branco, A. Pombo and M. P. Crippa (2010). "Poised transcription factories prime silent uPA gene prior to activation." PLoS Biol **8**(1): e1000270.

Flavahan, W. A., Y. Drier, B. B. Liau, S. M. Gillespie, A. S. Venteicher, A. O. Stemmer-Rachamimov, M. L. Suva and B. E. Bernstein (2016). "Insulator dysfunction and oncogene activation in IDH mutant gliomas." Nature **529**(7584): 110-114.

Flyamer, I. M., J. Gassler, M. Imakaev, H. B. Brandao, S. V. Ulianov, N. Abdennur, S. V. Razin, L. A. Mirny and K. Tachibana-Konwalski (2017). "Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition." Nature **544**(7648): 110-114.

Franke, M., D. M. Ibrahim, G. Andrey, W. Schwarzer, V. Heinrich, R. Schopflin, K. Kraft, R. Kempfer, I. Jerkovic, W. L. Chan, M. Spielmann, B. Timmermann, L. Wittler, I. Kurth, P. Cambiaso, O. Zuffardi, G. Houge, L. Lambie, F. Brancati, A. Pombo, M. Vingron, F. Spitz and S. Mundlos (2016). "Formation of new chromatin domains determines pathogenicity of genomic duplications." Nature **538**(7624): 265-269.

Franke, M. (2017). "The Role of Higher-Order Chromatin Organization at the SOX9 Locus in Gene Regulation and Disease." http://www.diss.fuberlin.de/diss/receive/FUDISS_thesis_000000105534

Fraser, J., C. Ferrai, A. M. Chiariello, M. Schueler, T. Rito, G. Laudanno, M. Barbieri, B. L. Moore, D. C. Kraemer, S. Aitken, S. Q. Xie, K. J. Morris, M. Itoh, H. Kawaji, I. Jaeger, Y. Hayashizaki, P. Carninci, A. R. Forrest, F. Consortium, C. A. Semple, J. Dostie, A. Pombo and M. Nicodemi (2015). "Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation." Mol Syst Biol **11**(12): 852.

Fudenberg, G. and M. Imakaev (2017). "FISH-ing for captured contacts: towards reconciling FISH and 3C." Nat Methods **14**(7): 673-678.

Fukaya, T., B. Lim and M. Levine (2016). "Enhancer Control of Transcriptional Bursting." Cell **166**(2): 358-368.

Gilbert, N. and J. Allan (2014). "Supercoiling in DNA and chromatin." Curr Opin Genet Dev **25**: 15-21.

Gilbert, N., S. Boyle, H. Fiegler, K. Woodfine, N. P. Carter and W. A. Bickmore (2004). "Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers." Cell **118**(5): 555-566.

Gilbert, N., S. Boyle, H. Sutherland, J. de Las Heras, J. Allan, T. Jenuwein and W. A. Bickmore (2003). "Formation of facultative heterochromatin in the absence of HP1." EMBO J **22**(20): 5540-5550.

Giorgetti, L., R. Galupa, E. P. Nora, T. Piolot, F. Lam, J. Dekker, G. Tiana and E. Heard (2014). "Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription." Cell **157**(4): 950-963.

Golkaram, M., J. Jang, S. Hellander, K. S. Kosik and L. R. Petzold (2017). "The Role of Chromatin Density in Cell Population Heterogeneity during Stem Cell Differentiation." Sci Rep **7**(1): 13307.

Gordon, C. T., T. Y. Tan, S. Benko, D. Fitzpatrick, S. Lyonnet and P. G. Farlie (2009). "Long-range regulation at the SOX9 locus in development and disease." J Med Genet **46**(10): 649-656.

Grasser, F., M. Neusser, H. Fiegler, T. Thormeyer, M. Cremer, N. P. Carter, T. Cremer and S. Muller (2008). "Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei." J Cell Sci **121**(11): 1876-1886.

Grosberg, A., S. K. Nechaev and E. I. Shakhnovich (1988). "[The role of topological limitations in the kinetics of homopolymer collapse and self-assembly of biopolymers]." Biofizika **33**(2): 247-253.

Guelen, L., L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat and B. van Steensel (2008). "Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions." Nature **453**(7197): 948-951.

Guillot, P. V., S. Q. Xie, M. Hollinshead and A. Pombo (2004). "Fixation-induced redistribution of hyperphosphorylated RNA polymerase II in the nucleus of human cells." Exp Cell Res **295**(2): 460-468.

Guo, Y., X. Gu, Z. Sheng, Y. Wang, C. Luo, R. Liu, H. Qu, D. Shu, J. Wen, R. P. Crooijmans, O. Carlborg, Y. Zhao, X. Hu and N. Li (2016). "A Complex Structural Variation on Chromosome 27 Leads to the Ectopic Expression of HOXB8 and the Muffs and Beard Phenotype in Chickens." PLoS Genet **12**(6): e1006071.

Hall, L. L., K. P. Smith, M. Byron and J. B. Lawrence (2006). "Molecular anatomy of a speckle." Anat Rec A Discov Mol Cell Evol Biol **288**(7): 664-675.

Hamdan, F. H. and S. A. Johnsen (2018). "Super enhancers - new analyses and perspectives on the low hanging fruit." Transcription **9**(2): 123-130.

Hay, D., J. R. Hughes, C. Babbs, J. O. J. Davies, B. J. Graham, L. Hanssen, M. T. Kassouf, A. M. Marieke Oudelaar, J. A. Sharpe, M. C. Suciu, J. Telenius, R. Williams, C. Rode, P. S. Li, L. A. Pennacchio, J. A. Sloane-Stanley, H. Ayyub, S. Butler, T. Sauka-Spengler, R. J. Gibbons, A. J. H. Smith, W. G. Wood and D. R. Higgs (2016). "Genetic dissection of the alpha-globin super-enhancer in vivo." Nat Genet **48**(8): 895-903.

Hnisz, D., B. J. Abraham, T. I. Lee, A. Lau, V. Saint-Andre, A. A. Sigova, H. A. Hoke and R. A. Young (2013). "Super-enhancers in the control of cell identity and disease." Cell **155**(4): 934-947.

Hnisz, D., D. S. Day and R. A. Young (2016). "Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control." Cell **167**(5): 1188-1200.

Hnisz, D., K. Shrinivas, R. A. Young, A. K. Chakraborty and P. A. Sharp (2017). "A Phase Separation Model for Transcriptional Control." Cell **169**(1): 13-23.

Hsin, J. P. and J. L. Manley (2012). "The RNA polymerase II CTD coordinates transcription and RNA processing." Genes Dev **26**(19): 2119-2137.

Ibarra, A., C. Benner, S. Tyagi, J. Cool and M. W. Hetzer (2016). "Nucleoporin-mediated regulation of cell identity genes." Genes Dev **30**(20): 2253-2258.

Iborra, F. J., A. Pombo, D. A. Jackson and P. R. Cook (1996). "Active RNA polymerases are localized within discrete transcription "factories' in human nuclei." J Cell Sci **109 (Pt 6)**: 1427-1436.

Iborra, F. J., D. A. Jackson and P. R. Cook (1998). "The path of transcripts from extra-nucleolar synthetic sites to nuclear pores: transcripts in transit are concentrated in discrete structures containing SR proteins." J Cell Sci **111 (Pt 15)**: 2269-2282.

Javierre, B. M., O. S. Burren, S. P. Wilder, R. Kreuzhuber, S. M. Hill, S. Sewitz, J. Cairns, S. W. Wingett, C. Varnai, M. J. Thiecke, F. Burden, S. Farrow, A. J. Cutler, K. Rehnstrom, K. Downes, L. Grassi, M. Kostadima, P. Freire-Pritchett, F. Wang, B. Consortium, H. G. Stunnenberg, J. A. Todd, D. R. Zerbino, O. Stegle, W. H. Ouwehand, M. Frontini, C. Wallace, M. Spivakov and P. Fraser (2016). "Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters." Cell **167**(5): 1369-1384 e1319.

Kashimada, K. and P. Koopman (2010). "Sry: the master switch in mammalian sex determination." Development **137**(23): 3921-3930.

Klopocki, E. and S. Mundlos (2011). "Copy-number variations, noncoding sequences, and human phenotypes." Annu Rev Genomics Hum Genet **12**: 53-72.

Koch, F., R. Fenouil, M. Gut, P. Cauchy, T. K. Albert, J. Zacarias-Cabeza, S. Spicuglia, A. L. de la Chapelle, M. Heidemann, C. Hintermair, D. Eick, I. Gut, P. Ferrier and J. C. Andrau (2011). "Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters." Nat Struct Mol Biol **18**(8): 956-963.

Kosak, S. T., J. A. Skok, K. L. Medina, R. Riblet, M. M. Le Beau, A. G. Fisher and H. Singh (2002). "Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development." Science **296**(5565): 158-162.

Kurth, I., E. Klopocki, S. Stricker, J. van Oosterwijk, S. Vanek, J. Altmann, H. G. Santos, J. J. van Harssel, T. de Ravel, A. O. Wilkie, A. Gal and S. Mundlos (2009). "Duplications of noncoding elements 5' of SOX9 are associated with brachydactyly-anonychia." Nat Genet **41**(8): 862-863.

Larson, A. G., D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding and G. J. Narlikar (2017). "Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin." Nature **547**(7662): 236-240.

Levings, P. P. and J. Bungert (2002). "The human beta-globin locus control region." Eur J Biochem **269**(6): 1589-1599.

Li, Y., L. M. Almassalha, J. E. Chandler, X. Zhou, Y. E. Stypula-Cyrus, K. A. Hujsak, E. W. Roth, R. Bleher, H. Subramanian, I. Szleifer, V. P. Dravid and V. Backman (2017). "The effects of chemical fixation on the cellular nanostructure." Exp Cell Res **358**(2): 253-259.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander and J. Dekker (2009). "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." Science **326**(5950): 289-293.

Loven, J., H. A. Hoke, C. Y. Lin, A. Lau, D. A. Orlando, C. R. Vakoc, J. E. Bradner, T. I. Lee and R. A. Young (2013). "Selective inhibition of tumor oncogenes by disruption of super-enhancers." Cell **153**(2): 320-334.

Lupianez, D. G., K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel and S. Mundlos (2015). "Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions." Cell **161**(5): 1012-1025.

Lybaek, H., D. de Bruijn, A. H. den Engelsman-van Dijk, D. Vanichkina, C. Nepal, A. Brendehaug and G. Houge (2014). "RevSex duplication-induced and sex-related differences in the SOX9 regulatory region chromatin landscape in human fibroblasts." Epigenetics **9**(3): 416-427.

Martin, S. and A. Pombo (2003). "Transcription factories: quantitative studies of nanostructures in the mammalian nucleus." Chromosome Res **11**(5): 461-470.

Mason, J. T. and T. J. O'Leary (1991). "Effects of formaldehyde fixation on protein secondary structure: a calorimetric and infrared spectroscopic investigation." J Histochem Cytochem **39**(2): 225-229.

McDowall, A., J. Gruenberg, K. Romisch and G. Griffiths (1989). "The structure of organelles of the endocytic pathway in hydrated cryosections of cultured cells." Eur J Cell Biol **49**(2): 281-294.

McKeown, P. C. and P. J. Shaw (2009). "Chromatin: linking structure and function in the nucleolus." Chromosoma **118**(1): 11-23.

Mefford, H. C., A. J. Sharp, C. Baker, A. Itsara, Z. Jiang, K. Buysse, S. Huang, V. K. Maloney, J. A. Crolla, D. Baralle, A. Collins, C. Mercer, K. Norga, T. de Ravel, K. Devriendt, E. M. Bongers, N. de Leeuw, W. Reardon, S. Gimelli, F. Bena, R. C. Hennekam, A. Male, L. Gaunt, J. Clayton-Smith, I. Simonic, S. M. Park, S. G. Mehta, S. Nik-Zainal, C. G. Woods, H. V. Firth, G. Parkin, M. Fichera, S. Reitano, M. Lo Giudice, K. E. Li, I. Casuga, A. Broomer, B. Conrad, M. Schwerzmann, L. Raber, S. Gallati, P. Striano, A. Coppola, J. L.

Tolmie, E. S. Tobias, C. Lilley, L. Armengol, Y. Spysschaert, P. Verloo, A. De Coene, L. Goossens, G. Mortier, F. Speleman, E. van Binsbergen, M. R. Nelen, R. Hochstenbach, M. Poot, L. Gallagher, M. Gill, J. McClellan, M. C. King, R. Regan, C. Skinner, R. E. Stevenson, S. E. Antonarakis, C. Chen, X. Estivill, B. Menten, G. Gimelli, S. Gribble, S. Schwartz, J. S. Sutcliffe, T. Walsh, S. J. Knight, J. Sebat, C. Romano, C. E. Schwartz, J. A. Veltman, B. B. de Vries, J. R. Vermeesch, J. C. Barber, L. Willatt, M. Tassabehji and E. E. Eichler (2008). "Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes." N Engl J Med **359**(16): 1685-1699.

Melnik, S., B. Deng, A. Papantonis, S. Baboo, I. M. Carr and P. R. Cook (2011). "The proteomes of transcription factories containing RNA polymerases I, II or III." Nat Methods **8**(11): 963-968.

Mifsud, B., F. Tavares-Cadete, A. N. Young, R. Sugar, S. Schoenfelder, L. Ferreira, S. W. Wingett, S. Andrews, W. Grey, P. A. Ewels, B. Herman, S. Happe, A. Higgs, E. LeProust, G. A. Follows, P. Fraser, N. M. Luscombe and C. S. Osborne (2015). "Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C." Nat Genet **47**(6): 598-606.

Mongelard, F., C. Vourc'h, M. Robert-Nicoud and Y. Usson (1999). "Quantitative assessment of the alteration of chromatin during the course of FISH procedures. Fluorescent in situ hybridization." Cytometry **36**(2): 96-101.

Moorthy, S. D., S. Davidson, V. M. Shchuka, G. Singh, N. Malek-Gilani, L. Langroudi, A. Martchenko, V. So, N. N. Macpherson and J. A. Mitchell (2017). "Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes." Genome Res **27**(2): 246-258.

Moorthy, S. D. and J. A. Mitchell (2016). "Generating CRISPR/Cas9 Mediated Monoallelic Deletions to Study Enhancer Function in Mouse Embryonic Stem Cells." J Vis Exp(110): e53552.

Muller, T., C. Schumann and A. Kraegeloh (2012). "STED microscopy and its applications: new insights into cellular processes on the nanoscale." Chemphyschem **13**(8): 1986-2000.
Torok, P. and P. Munro (2004). "The use of Gauss-Laguerre vector beams in STED microscopy." Opt Express **12**(15): 3605-3617.

Narendra, V., M. Bulajic, J. Dekker, E. O. Mazzoni and D. Reinberg (2016). "CTCF-mediated topological boundaries during development foster appropriate gene regulation." Genes Dev **30**(24): 2657-2662.

Nemeth, A., A. Conesa, J. Santoyo-Lopez, I. Medina, D. Montaner, B. Peterfia, I. Solovei, T. Cremer, J. Dopazo and G. Langst (2010). "Initial genomics of the human nucleolus." PLoS Genet **6**(3): e1000889.

Ng, L. J., S. Wheatley, G. E. Muscat, J. Conway-Campbell, J. Bowles, E. Wright, D. M. Bell, P. P. Tam, K. S. Cheah and P. Koopman (1997). "SOX9 binds DNA, activates transcription, and coexpresses with type II collagen during chondrogenesis in the mouse." Dev Biol **183**(1): 108-121.

Nagano, T., Y. Lubling, C. Varnai, C. Dudley, W. Leung, Y. Baran, N. Mendelson Cohen, S. Wingett, P. Fraser and A. Tanay (2017). "Cell-cycle dynamics of chromosomal organization at single-cell resolution." Nature **547**(7661): 61-67.

Nagano, T., S. W. Wingett and P. Fraser (2017). "Capturing Three-Dimensional Genome Organization in Individual Cells by Single-Cell Hi-C." Methods Mol Biol **1654**: 79-97.

Nagy, K. and Nichols, J. (2011). "Derivation of Murine ES Cell Line". Advanced Protocols for Animal Transgenesis, Springer Protocols. 432-455.

Noordermeer, D., M. R. Branco, E. Splinter, P. Klous, W. van Ijcken, S. Swagemakers, M. Koutsourakis, P. van der Spek, A. Pombo and W. de Laat (2008). "Transcription and chromatin organization of a housekeeping gene cluster containing an integrated beta-globin locus control region." PLoS Genet **4**(3): e1000016.

Nora, E. P., B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Bluthgen, J. Dekker and E. Heard (2012). "Spatial partitioning of the regulatory landscape of the X-inactivation centre." Nature **485**(7398): 381-385.

Pennacchio, L. A., W. Bickmore, A. Dean, M. A. Nobrega and G. Bejerano (2013). "Enhancers: five essential questions." Nat Rev Genet **14**(4): 288-295.

Peric-Hupkes, D., W. Meuleman, L. Pagie, S. W. Bruggeman, I. Solovei, W. Brugman, S. Graf, P. Flicek, R. M. Kerkhoven, M. van Lohuizen, M. Reinders, L. Wessels and B. van Steensel (2010). "Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation." Mol Cell **38**(4): 603-613.

Phillips-Cremins, J. E., M. E. Sauria, A. Sanyal, T. I. Gerasimova, B. R. Lajoie, J. S. Bell, C. T. Ong, T. A. Hookway, C. Guo, Y. Sun, M. J. Bland, W. Wagstaff, S. Dalton, T. C. McDevitt, R. Sen, J. Dekker, J. Taylor and V. G. Corces (2013). "Architectural protein subclasses shape 3D organization of genomes during lineage commitment." Cell **153**(6): 1281-1295.

Pombo, A. (2007). "Advances in imaging the interphase nucleus using thin cryosections." Histochem Cell Biol **128**(2): 97-104.

Pombo, A. and M. R. Branco (2007). "Functional organisation of the genome during interphase." Curr Opin Genet Dev **17**(5): 451-455.

Pombo, A. and N. Dillon (2015). "Three-dimensional genome architecture: players and mechanisms." Nat Rev Mol Cell Biol **16**(4): 245-257.

Pope, B. D., T. Ryba, V. Dileep, F. Yue, W. Wu, O. Denas, D. L. Vera, Y. Wang, R. S. Hansen, T. K. Canfield, R. E. Thurman, Y. Cheng, G. Gulsoy, J. H. Dennis, M. P. Snyder, J. A. Stamatoyannopoulos, J. Taylor, R. C. Hardison, T. Kahveci, B. Ren and D. M. Gilbert (2014). "Topologically associating domains are stable units of replication-timing regulation." Nature **515**(7527): 402-405.

Preiss, S., A. Argentaro, A. Clayton, A. John, D. A. Jans, T. Ogata, T. Nagai, I. Barroso, A. J. Schafer and V. R. Harley (2001). "Compound effects of point mutations causing campomelic

dysplasia/autosomal sex reversal upon SOX9 structure, nuclear transport, DNA binding, and transcriptional activation." J Biol Chem **276**(30): 27864-27872.

Prokocimer, M., M. Davidovich, M. Nissim-Rafinia, N. Wiesel-Motiuk, D. Z. Bar, R. Barkan, E. Meshorer and Y. Gruenbaum (2009). "Nuclear lamins: key regulators of nuclear structure and activities." J Cell Mol Med **13**(6): 1059-1085.

Puvion, E. and F. Puvion-Dutilleul (1996). "Ultrastructure of the nucleus in relation to transcription and splicing: roles of perichromatin fibrils and interchromatin granules." Exp Cell Res **229**(2): 217-225.

Schindelin, J., Arganda-Carreras, I. and Frise, E. et al. (2012). "Fiji: an open-source platform for biological-image analysis", Nature methods **9**(7): 676-682.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012), "NIH Image to ImageJ: 25 years of image analysis", Nature methods **9**(7): 671-675.

Sofia A Quinodoz, Noah Ollikainen, Barbara Tabak, Ali Palla, Jan Marten Schmidt, Elizabeth Detmar, Mason Lai, Alexander Shishkin, Prashant Bhat, Vickie Trinh, Erik Aznauryan, Pamela Russell, Christine Cheng, Marko Jovanovic, Amy Chow, Patrick McDonel, Manuel Garber, Mitchell Guttman (2017). "Higher-order inter-chromosomal hubs shape 3-dimensional genome organization in the nucleus". bioRxiv
doi: https://doi.org/10.1101/219683

Rao, S. S., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander and E. L. Aiden (2014). "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." Cell **159**(7): 1665-1680.

Rao, S. S. P., S. C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, K. R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander and E. L. Aiden (2017). "Cohesin Loss Eliminates All Loop Domains." Cell **171**(2): 305-320 e324.

Schmitt, A. D., M. Hu, I. Jung, Z. Xu, Y. Qiu, C. L. Tan, Y. Li, S. Lin, Y. Lin, C. L. Barr and B. Ren (2016). "A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome." Cell Rep **17**(8): 2042-2059.

Schwarzer, W., N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie, N. A. Fonseca, W. Huber, H. H. C, L. Mirny and F. Spitz (2017). "Two independent modes of chromatin organization revealed by cohesin removal." Nature **551**(7678): 51-56.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T. C. Gilliam, B. Trask, N. Patterson, A. Zetterberg and M. Wigler (2004). "Large-scale copy number polymorphism in the human genome." Science **305**(5683): 525-528.

Sewitz, S. A., Z. Fahmi and K. Lipkow (2017). "Higher order assembly: folding the chromosome." Curr Opin Struct Biol **42**: 162-168.

Sexton, T., E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay and G. Cavalli (2012). "Three-dimensional folding and functional organization principles of the Drosophila genome." Cell **148**(3): 458-472.

Sharp, A. J., S. Hansen, R. R. Selzer, Z. Cheng, R. Regan, J. A. Hurst, H. Stewart, S. M. Price, E. Blair, R. C. Hennekam, C. A. Fitzpatrick, R. Segraves, T. A. Richmond, C. Guiver, D. G. Albertson, D. Pinkel, P. S. Eis, S. Schwartz, S. J. Knight and E. E. Eichler (2006). "Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome." Nat Genet **38**(9): 1038-1042.

Shopland, L. S., C. V. Johnson, M. Byron, J. McNeil and J. B. Lawrence (2003). "Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods." J Cell Biol **162**(6): 981-990.

Small, S., A. Blair and M. Levine (1992). "Regulation of even-skipped stripe 2 in the Drosophila embryo." EMBO J **11**(11): 4047-4057.

Smith, E. M., B. R. Lajoie, G. Jain and J. Dekker (2016). "Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus." Am J Hum Genet **98**(1): 185-201.

Solovei, I., A. Cavallo, L. Schermelleh, F. Jaunin, C. Scasselati, D. Cmarko, C. Cremer, S. Fakan and T. Cremer (2002). "Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence in situ hybridization (3D-FISH)." Exp Cell Res **276**(1): 10-23.

Solovei, I., M. Kreysing, C. Lanctot, S. Kosem, L. Peichl, T. Cremer, J. Guck and B. Joffe (2009). "Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution." Cell **137**(2): 356-368.

Spitz, F. and E. E. Furlong (2012). "Transcription factors: from enhancer binding to developmental control." Nat Rev Genet **13**(9): 613-626.

Strom, A. R., A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq and G. H. Karpen (2017). "Phase separation drives heterochromatin domain formation." Nature **547**(7662): 241-245.

Tanabe, H., S. Muller, M. Neusser, J. von Hase, E. Calcagno, M. Cremer, I. Solovei, C. Cremer and T. Cremer (2002). "Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates." Proc Natl Acad Sci U S A **99**(7): 4424-4429.

Thavarajah, R., V. K. Mudimbaimannar, J. Elizabeth, U. K. Rao and K. Ranganathan (2012). "Chemical and physical basics of routine formaldehyde fixation." J Oral Maxillofac Pathol **16**(3): 400-405.

Thibodeau, A., E. J. Marquez, D. G. Shin, P. Vera-Licona and D. Ucar (2017). "Chromatin interaction networks revealed unique connectivity patterns of broad H3K4me3 domains and super enhancers in 3D chromatin." Sci Rep **7**(1): 14466.

Tolhuis, B., R. J. Palstra, E. Splinter, F. Grosveld and W. de Laat (2002). "Looping and interaction between hypersensitive sites in the active beta-globin locus." Mol Cell **10**(6): 1453-1465.

Tokuyasu, K. T. (1973). "A technique for ultracryotomy of cell suspensions and tissues." J Cell Biol **57**(2): 551-565.

Tokuyasu, K. T. (1980). "Immunochemistry on ultrathin frozen sections." Histochem J **12**(4): 381-403.

Torok, P. and P. Munro (2004). "The use of Gauss-Laguerre vector beams in STED microscopy." Opt Express **12**(15): 3605-3617.

van Koningsbruggen, S., M. Gierlinski, P. Schofield, D. Martin, G. J. Barton, Y. Ariyurek, J. T. den Dunnen and A. I. Lamond (2010). "High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli." Mol Biol Cell **21**(21): 3735-3748.

van Steensel, B. and A. S. Belmont (2017). "Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression." Cell **169**(5): 780-791.

Wang, S., J. H. Su, B. J. Beliveau, B. Bintu, J. R. Moffitt, C. T. Wu and X. Zhuang (2016). "Spatial organization of chromatin domains and compartments in single chromosomes." Science **353**(6299): 598-602.

Weintraub, A. S., C. H. Li, A. V. Zamudio, A. A. Sigova, N. M. Hannett, D. S. Day, B. J. Abraham, M. A. Cohen, B. Nabet, D. L. Buckley, Y. E. Guo, D. Hnisz, R. Jaenisch, J. E. Bradner, N. S. Gray and R. A. Young (2017). "YY1 Is a Structural Regulator of Enhancer-Promoter Loops." Cell **171**(7): 1573-1588 e1528.

Westphal, V., S. O. Rizzoli, M. A. Lauterbach, D. Kamin, R. Jahn and S. W. Hell (2008). "Video-rate far-field optical nanoscopy dissects synaptic vesicle movement." Science **320**(5873): 246-249.

Whyte, W. A., D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee and R. A. Young (2013). "Master transcription factors and mediator establish super-enhancers at key cell identity genes." Cell **153**(2): 307-319.

Williams, R. R., V. Azuara, P. Perry, S. Sauer, M. Dvorkina, H. Jorgensen, J. Roix, P. McQueen, T. Misteli, M. Merkenschlager and A. G. Fisher (2006). "Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus." J Cell Sci **119**(Pt 1): 132-140.

Williamson, I., S. Berlivet, R. Eskeland, S. Boyle, R. S. Illingworth, D. Paquette, J. Dostie and W. A. Bickmore (2014). "Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization." Genes Dev **28**(24): 2778-2791.

Xie, S. Q., S. Martin, P. V. Guillot, D. L. Bentley and A. Pombo (2006). "Splicing speckles are not reservoirs of RNA polymerase II, but contain an inactive form, phosphorylated on serine2 residues of the C-terminal domain." Mol Biol Cell **17**(4): 1723-1733.

Yao, B., Q. Wang, C. F. Liu, P. Bhattaram, W. Li, T. J. Mead, J. F. Crish and V. Lefebvre (2015). "The SOX9 upstream region prone to chromosomal aberrations causing campomelic dysplasia contains multiple cartilage enhancers." Nucleic Acids Res **43**(11): 5394-5408.

Zarrei, M., J. R. MacDonald, D. Merico and S. W. Scherer (2015). "A copy number variation map of the human genome." Nat Rev Genet **16**(3): 172-183.

Zhao, Q., H. Eberspaecher, V. Lefebvre and B. De Crombrugghe (1997). "Parallel expression of Sox9 and Col2a1 in cells undergoing chondrogenesis." Dev Dyn **209**(4): 377-386.

7 Appendix

7.1 3D Image analysis pipeline

The matlab script was devised by Till Hülnhagen and Dorothee Krämer, written by Till Hülnhagen and edited by Dorothee Krämer.

```
%% Script to segment and analyze FISH signal in multi-channel microscopy images

% 2017-07-18, Till Huelnhagen, created

    close all; clear; clc;

    addpath(genpath('/Users/Dorothee/Documents/MATLAB'))

    % define colormaps
    cmapR=gray(256); cmapR(:,[2,3])=0;
    cmapG=gray(256); cmapG(:,[1,3])=0;
    cmapB=gray(256); cmapB(:,[1,2])=0;

    % cmap = colourmap
%% load data from disk
    datasetNum=1;
    dataPath=sprintf('/Users/Dorothee/Desktop/2016-01-11 Sox9Dublication Sox9 probes, seperate
channels/Stack,%d',datasetNum);
    datapathG=fullfile(dataPath,'ch00');
    datapathR=fullfile(dataPath,'ch01');
    datapathB=fullfile(dataPath,'ch02');
    datapathBG=fullfile(dataPath,'ch03');

   % define Voxel dimensions
    voxx=90
    voxy=90
    voxz=200

    % green
    filesG=dir(fullfile(datapathG,'*.tif'));
    imgG=imread(fullfile(datapathG,filesG(1).name));
    imgG=zeros([size(imgG),numel(filesG)],'single');
    for file = 1:numel(filesG)
        imgG(:,:,:,file) = imread(fullfile(datapathG,filesG(file).name));
    end
    % select only green channel and scale to [0 1]
    imgG=squeeze(imgG(:,:,2,:)./255);

    % red
    filesR=dir(fullfile(datapathR,'*.tif'));
    imgR=imread(fullfile(datapathR,filesR(1).name));
    imgR=zeros([size(imgR),numel(filesR)],'single');
    for file = 1:numel(filesR)
        imgR(:,:,:,file) = imread(fullfile(datapathR,filesR(file).name));
```

```
    end
    % select only red channel and scale to [0 1]
    imgR=squeeze(imgR(:,:,1,:)./255);

    % blue
    filesB=dir(fullfile(datapathB,'*.tif'));
    imgB=imread(fullfile(datapathB,filesB(1).name));
    imgB=zeros([size(imgB),numel(filesB)],'single');
    for file = 1:numel(filesB)
        imgB(:,:,:,file) = imread(fullfile(datapathB,filesB(file).name));
    end
    % select only blue channel and scale to [0 1]
    imgB=squeeze(imgB(:,:,3,:)./255);

    % Background
    filesBG=dir(fullfile(datapathBG,'*.tif'));
    imgBG=imread(fullfile(datapathBG,filesBG(1).name));
    imgBG=zeros([size(imgBG),numel(filesBG)],'single');
    for file = 1:numel(filesBG)
        imgBG(:,:,:,file) = imread(fullfile(datapathBG,filesBG(file).name));
    end
    % select only one channel and scale to [0 1]
    imgBG=squeeze(imgBG(:,:,1,:)./255);

%% display green image to test
    imagesc(medfilt2(squeeze(imgG(:,:,18)),[3 3]),[0.05 0.1]); colormap(cmapG); axis image;

%% display RGB channels combined
    sl=16;
    imgTmp=cat(3,imgR(:,:,sl),imgG(:,:,sl),imgB(:,:,sl));
    %imgTmp=cat(3,medfilt2(imgR(:,:,sl),[3 3]),medfilt2(imgG(:,:,sl),[3 3]),medfilt2(imgB(:,:,sl),[3 3]));
    % adjust brightness

    imgTmp=imgTmp.*3;
    imgTmp=imgTmp + repmat(imgBG(:,:,sl),[1 1 3])*3;
    imgTmp(imgTmp>1)=1;
    imagesc(imgTmp); axis image; shg;
    clear imgTmp;

    %% create RBG image of background and color channels
    brightnessFactor=3;
    imgRGB=cat(4,imgR,imgG,imgB) + repmat(imgBG,[1 1 1 3]);
    imgRGB=permute(imgRGB,[1 2 4 3]);
    % adjust brightness
    imgRGB=imgRGB.*brightnessFactor;
    imgRGB(imgRGB>1)=1;

  %% thresholding of red channel
%     imcontrast
%     imcontrast(h)
%     hfigure = imcontrast(...)
%
%
    %% find threshold for red
    sl=13;
    threshR=zeros(8);
    %threshR(1)=0.22;
    threshR(1)=0.15;
    threshR(2)=0.17;
```

```
    threshR(3)=0.12;
    threshR(4)=0.12;
    threshR(5)=0.15;
    threshR(6)=0.16;
    threshR(7)=0.1;
    threshR(8)=0.10;
    imgRThresh=imgR;
    imgRThresh(imgRThresh<threshR(datasetNum))=0;
    subplot(121)
        imagesc(imgRGB(:,:,:,sl)); axis image;
        title('original image');
    subplot(122)
        imagesc(imgRThresh(:,:,sl),[0 0.1]); axis image; colormap(cmapR);

%% find threshold for green
    sl=26;
    threshG=zeros(8);
    threshG(1)=0.10;
    threshG(2)=0.02;
    threshG(3)=0.06;
    threshG(4)=0.06;
    threshG(5)=0.00;
    threshG(6)=0;
    threshG(7)=0;
    threshG(8)=0.07;
    imgGThresh=imgG;
    imgGThresh(imgGThresh<threshG(datasetNum))=0;
    subplot(121)
        imagesc(imgRGB(:,:,:,sl)); axis image;
        title('original image');
    subplot(122)
        imagesc(imgGThresh(:,:,sl),[0 0.1]); axis image; colormap(cmapG);

%% find threshold for blue
    sl=34;
    threshB=zeros(8);
    threshB(1)=0.15;
    threshB(2)=0;
    threshB(3)=0.10;
    threshB(4)=0.14;
    threshB(5)=0.02;
    threshB(6)=0.02;
    threshB(7)=0.02;
    threshB(8)=0.07;
    imgBThresh=imgB;
    imgBThresh(imgBThresh<threshB(datasetNum))=0;
    subplot(121)
        imagesc(imgRGB(:,:,:,sl)); axis image;
        title('original image');
    subplot(122)
        imagesc(imgBThresh(:,:,sl),[0 0.1]); axis image; colormap(cmapB);

%% apply threshold and 3D median filter to get rid of background
%     threshG=0.12;
%     threshR=0.11;
%     threshB=0.19;

  medianFilterKernel=[3,3,3];
```

```
    imgGThresh=imgG;
    imgGThresh(imgGThresh<threshG(datasetNum))=0;
    imgGThreshFilt=medfilt3(imgGThresh,medianFilterKernel);
    imGFilt=medfilt3(imgG,medianFilterKernel);

    imgRThresh=imgR;
    imgRThresh(imgRThresh<threshR(datasetNum))=0;
    imgRThreshFilt=medfilt3(imgRThresh,medianFilterKernel);
    imRFilt=medfilt3(imgR,medianFilterKernel);

    imgBThresh=imgB;
    imgBThresh(imgBThresh<threshB(datasetNum))=0;
    imgBThreshFilt=medfilt3(imgBThresh,medianFilterKernel);
    imBFilt=medfilt3(imgB,medianFilterKernel);

%% display filtering result for green channel
    sl=34
    subplot(221)
        imagesc(imgG(:,:,sl),[0 0.1]); axis image;
        title('original image');

    subplot(222)
        imagesc(imgGThresh(:,:,sl),[0 0.1]); axis image;
        title('thresholded image');
    subplot(223)
        imagesc(imgGThreshFilt(:,:,sl),[0.0 0.1]); axis image;
        title('filtereded thresholded image');
    subplot(224)
        imagesc(imGFilt(:,:,sl),[0.03 0.1]); colormap(cmapG); axis image;
        title('thresholded filtereded image'); shg;

%% display filtering result for red channel
    sl=32;
    subplot(221)
        imagesc(imgR(:,:,sl),[0 0.1]); axis image;
        title('original image');
    subplot(222)
        imagesc(imgRThresh(:,:,sl),[0 0.1]); axis image;
        title('thresholded image');
    subplot(223)
        imagesc(imgRThreshFilt(:,:,sl),[0.0 0.1]); axis image;
        title('filtereded thresholded image');
    subplot(224)
        imagesc(imRFilt(:,:,sl),[0.03 0.1]); colormap(cmapR); axis image;
        title('thresholded filtereded image'); shg;

%% display filtering result for blue channel
    sl=29;
    subplot(221)
        imagesc(imgB(:,:,sl),[0 0.1]); axis image;
        title('original image');
    subplot(222)
        imagesc(imgBThresh(:,:,sl),[0 0.1]); axis image;
        title('thresholded image');
    subplot(223)
        imagesc(imgBThreshFilt(:,:,sl),[0.0 0.1]); axis image;
        title('filtereded thresholded image');
    subplot(224)
        imagesc(imBFilt(:,:,sl),[0.05 0.1]); colormap(cmapB); axis image;
```

```
    title('thresholded filtereded image'); shg;

%% create RBG image with background and all color channels for comparison
   brightnessFactor=3;
   imgRGB=cat(4,imgR,imgG,imgB) + repmat(imgBG,[1 1 1 3]);
   imgRGB=permute(imgRGB,[1 2 4 3]);
   % adjust brightness
   imgRGB=imgRGB.*brightnessFactor;
   imgRGB(imgRGB>1)=1;

   imgRGBThreshFilt=cat(4,imgRThreshFilt,imgGThreshFilt,imgBThreshFilt);
   imgRGBThreshFilt(isnan(imgRGBThreshFilt))=0;
   imgRGBThreshFilt=imgRGBThreshFilt + repmat(imgBG,[1 1 1 3]);
   imgRGBThreshFilt=permute(imgRGBThreshFilt,[1 2 4 3]);
   % adjust brightness
   imgRGBThreshFilt=imgRGBThreshFilt.*brightnessFactor;
   imgRGBThreshFilt(imgRGBThreshFilt>1)=1;

   %% display combined image pre and post filtering together with segmented images
   sl=7;
   subplot(231)
      imagesc(imgRGB(:,:,:,sl)); axis image;
   subplot(232)
      imagesc(imgRGBThreshFilt(:,:,:,sl)); axis image;
   subplot(234)

imagesc(cat(3,single(imgRThreshFilt(:,:,sl)>0),zeros(size(imgR(:,:,sl)),'single'),zeros(size(imgR(:,:,sl)),'
single'))); axis image;
   subplot(235)

imagesc(cat(3,zeros(size(imgG(:,:,sl)),'single'),single(imgGThreshFilt(:,:,sl)>0),zeros(size(imgG(:,:,sl)),
'single'))); axis image;
   subplot(236)

imagesc(cat(3,zeros(size(imgB(:,:,sl)),'single'),zeros(size(imgB(:,:,sl)),'single'),single(imgBThreshFilt(:,
:,sl)>0))); axis image; shg;

%% perform region property analysis

   % label regions
   maskG=(imgGThreshFilt>0);
   [lmG, nG] = bwlabeln(squeeze(maskG), 26);
   statsG = regionprops(lmG, {'Area','Centroid'});

%     remove regions with area < 10 or area > 1000
%     for i = 1:numel(statsG)
%        if (statsG(i,1).Area<10 || statsG(i,1).Area>2000), lmG(lmG==i)=0; end
%     end
%     clear i
%



   % remove regions with area < 80
   for i = 1:numel(statsG)
      if statsG(i,1).Area<80, lmG(lmG==i)=0; end
   end
   clear i
```

```
    % relabel & take voxel dimensions into account
    [lmG, nG] = bwlabeln(lmG>0, 26);
    statsG = regionprops(lmG, {'Area','Centroid'});
    c=struct2cell(statsG);
    centroidG=cell2mat(c(2,:));
    centroidG=reshape(centroidG,[3,nG]);
    areaG=(cell2mat(c(1,:))*voxx*voxy*voxz);
    clear c;

% red
    % label regions
    maskR=(imgRThreshFilt>0);
    [lmR, nR] = bwlabeln(squeeze(maskR), 26);
    statsR = regionprops(lmR, {'Area','Centroid'});

%     remove regions with area < 20
    for i = 1:numel(statsR)
        if statsR(i,1).Area<20, lmR(lmR==i)=0; end
    end
    clear i

    % relabel & take voxel dimensions into account
    [lmR, nR] = bwlabeln(lmR>0, 26);
    statsR = regionprops(lmR, {'Area','Centroid'});
    c=struct2cell(statsR);
    centroidR=cell2mat(c(2,:));
    centroidR=reshape(centroidR,[3,nR]);
    areaR=(cell2mat(c(1,:))*voxx*voxy*voxz);
    clear c;

% blue
    % label regions
    maskB=(imgBThreshFilt)>0;
    [lmB, nB] = bwlabeln(squeeze(maskB), 26);
    statsB = regionprops(lmB, {'Area','Centroid'});

    % remove regions with area < 32
    for i = 1:numel(statsB)
        if statsB(i,1).Area<32, lmB(lmB==i)=0; end
    end
    clear i



    % relabel & take voxel dimensions into account
    [lmB, nB] = bwlabeln(lmB>0, 26);
    statsB = regionprops(lmB, {'Area','Centroid'});
    c=struct2cell(statsB);
    centroidB=cell2mat(c(2,:));
    centroidB=reshape(centroidB,[3,nB]);
    areaB=(cell2mat(c(1,:))*voxx*voxy*voxz);
    clear c;

    lmR=int16(lmR);
    lmG=int16(lmG);
    lmB=int16(lmB);

%% Write Area in datatable
```

```
   dlmwrite('areaG.txt', areaG)
   dlmwrite('areaR.txt', areaR)
   dlmwrite('areaB.txt', areaB)

%% display label map and mask
   sl=30;
   subplot(121)
      imagesc(maskR(:,:,sl)); axis image;
      title('mask');
   subplot(122)
      imagesc(lmR(:,:,sl),[30 40]); axis image; colormap(cmapR);
      title('label map'); shg;


%% display combined image pre and post segmentation
   sl=26;
   subplot(231)
      imagesc(imgRGB(:,:,:,sl)); axis image;axis off;
   subplot(232)
      imagesc(imgRGBThreshFilt(:,:,:,sl)); axis image; axis off;
   subplot(234)
      imagesc(cat(3,double(lmR(:,:,sl)>0),zeros(size(imgR(:,:,sl))),zeros(size(imgR(:,:,sl))))); axis
image;axis off;
   subplot(235)
      imagesc(cat(3,zeros(size(imgG(:,:,sl))),double(lmG(:,:,sl)>0),zeros(size(imgG(:,:,sl))))); axis
image;axis off;
   subplot(236)
      imagesc(cat(3,zeros(size(imgB(:,:,sl))),zeros(size(imgB(:,:,sl))),double(lmB(:,:,sl)>0))); axis
image; axis off;shg;


%% plot histograms of blob sizes
   %bins=25:50:5000;
   h=bar(bins,[hist(areaR,bins);hist(areaG,bins);hist(areaB,bins)]');
%     xlim([0,5100]);
%     ylim([0,20]);
   title('Blob area');
   set(h(1),'FaceColor','r','EdgeColor','r');
   set(h(2),'FaceColor','g','EdgeColor','g');
   set(h(3),'FaceColor','b','EdgeColor','b');
   shg;

%% save masks as Nifti for rendering in imageJ
%     outPath='D:\studies\DK_data_analysis\segmentation';
%     save_nii(make_nii(squeeze(int16(lmR>0))),fullfile(outPath,'maskR.nii.gz'));
%     save_nii(make_nii(squeeze(int16(lmG>0))),fullfile(outPath,'maskG.nii.gz'));
%     save_nii(make_nii(squeeze(int16(lmB>0))),fullfile(outPath,'maskB.nii.gz'));

%% calculate distances between centroids


   scale=[voxx;voxy;voxz];

   nRed=size(centroidR,2);
   nGreen=size(centroidG,2);
   nBlue=size(centroidB,2);
```

```
% red to green
   distancesRG=zeros(nRed,nGreen,'single');
   for i = 1:nRed
      for j = 1:nGreen
         % distancesRG(i,j)=norm(centroidR(:,i)-centroidG(:,j));
         distancesRG(i,j)=norm((centroidR(:,i)-centroidG(:,j)).*scale);
      end
   end

% red to blue
   distancesRB=zeros(nRed,nBlue,'single');
   for i = 1:nRed
      for j = 1:nBlue
         % distancesRB(i,j)=norm(centroidR(:,i)-centroidB(:,j));
         distancesRB(i,j)=norm((centroidR(:,i)-centroidB(:,j)).*scale);
      end
   end

% green to blue
   distancesGB=zeros(nGreen,nBlue,'single');
   for i = 1:nGreen
      for j = 1:nBlue
         % distancesGB(i,j)=norm(centroidG(:,i)-centroidB(:,j));
         distancesGB(i,j)=norm((centroidG(:,i)-centroidB(:,j)).*scale);
      end
   end

%% plot histograms of distances
   subplot(131)
      hist(distancesRG(:))
      title('Distances red to green')
   subplot(132)
      hist(distancesRB(:))
      title('Distances red to blue')
   subplot(133)
      hist(distancesGB(:))
      title('Distances green to blue')

%% plot histogram of distances in one plot
   bins=25:50:600;
   h=bar(bins,[hist(distancesRG(:),bins);hist(distancesRB(:),bins);hist(distancesGB(:),bins)]');
   xlim([0,600]);
   title('centroid distances');
   set(h(1),'FaceColor','r','EdgeColor','g','LineWidth',4);
   set(h(2),'FaceColor','r','EdgeColor','b','LineWidth',4);
   set(h(3),'FaceColor','g','EdgeColor','b','LineWidth',4);
   legend('red to green','red to blue','green to blue');
   shg;

%% calculate distances to nearest blob of other color
   % red to green
   closestDistancesRG=min(distancesRG,[],2);
   % green to red
   closestDistancesGR=min(distancesRG,[],1);

   % red to blue
   closestDistancesRB=min(distancesRB,[],2);
   % blue to red
   closestDistancesBR=min(distancesRB,[],1);
```

```
    % green to blue
    closestDistancesGB=min(distancesGB,[],2);
    % blue to green
    closestDistancesBG=min(distancesGB,[],1);

    %% Write Area in datatable
    dlmwrite('closestneighbourCM_DistancesGR.txt', closestDistancesGR)
    dlmwrite('closestneighbourCM_DistancesBR.txt', closestDistancesBR)
    dlmwrite('closestneighbourCM_DistancesBG.txt', closestDistancesBG)

%% plot histograms of closest neighbor distances

    subplot(131)
    hist(closestDistancesRG(:));
        ylim([0,20]);
        xlim([0,40]);
        title('Closest distances red to green')
    subplot(132)
        hist(closestDistancesRB(:))
        ylim([0,20]);
        xlim([0,40]);
        title('Closest distances red to blue')
    subplot(133)
        hist(closestDistancesGB(:))
        ylim([0,20]);
        xlim([0,40]);
        title('Closes distances green to blue')

%% calculate border to border distances

    maskR=lmR>0;
    maskG=lmG>0;
    maskB=lmB>0;
    % remove inner part of blobs to save time
    %SE=strel('ball',3,3,0);
    SE=createSphereInVolume(3,3,3,1.5,1,0);
    shellR=int16(maskR-imerode(maskR,SE));
    shellG=int16(maskG-imerode(maskG,SE));
    shellB=int16(maskB-imerode(maskB,SE));

    % create labeled shells
    lShellR=lmR.*shellR;
    lShellG=lmG.*shellG;
    lShellB=lmB.*shellB;

    %% display shell
    imagesc(lShellR(:,:,20)); axis image;

    %% loop through all the voxels of each region and calculate the closest distance to a voxel of the
other regions

    %% red to green
    %[X,Y,Z]=meshgrid(1:size(imgG,1),1:size(imgG,2),1:size(imgG,3));
    [X,Y,Z]=meshgrid(0:voxx:(size(imgG,1)-1)*voxx,0:voxy:(size(imgG,2)-1)*voxy,0:voxz:(size(imgG,3)-
1)*voxz);
```

```
borderDistRG=nan(nR,nG);
   tic
   for i=1:nR
      for j=1:nG
         % check if regions are overlapping
         reg1=lmR==i;
         reg2=lmG==j;
         if sum2(and(reg1,reg2))>0
            minDist=0;
            clear reg1 reg2;
         else
            % get voxels belonging to the shells of the current regions
            indR = find(lShellR==i);
            indG = find(lShellG==j);
            % calculate distances of each red shell voxel to each green shell voxel
            minDist=inf;
            for k=1:numel(indR)
               for l=1:numel(indG)
                  dist=sqrt( (X(indR(k)) - X(indG(l))).^2 + (Y(indR(k)) - Y(indG(l))).^2 + (Z(indR(k)) -
Z(indG(l))).^2 );
                  if dist < minDist, minDist = dist; end
               end
            end
         end
         borderDistRG(i,j)=minDist;
      end
   end
   toc
   clear indR indG dist minDist X Y Z;


   %% red to blue
   %[X,Y,Z]=meshgrid(1:size(imgG,1),1:size(imgG,2),1:size(imgG,3));
   [X,Y,Z]=meshgrid(0:voxx:(size(imgG,1)-1)*voxx,0:voxy:(size(imgG,2)-1)*voxy,0:voxz:(size(imgG,3)-
1)*voxz);

   borderDistRB=nan(nR,nB);
   tic
   for i=1:nR
      for j=1:nB
         % check if regions are overlapping
         reg1=lmR==i;
         reg2=lmB==j;
         if sum2(and(reg1,reg2))>0
            minDist=0;
            clear reg1 reg2;
         else
            % get voxels belonging to the shells of the current regions
            indR = find(lShellR==i);
            indB = find(lShellB==j);
            % calculate distances of each red shell voxel to each blue shell voxel
            minDist=inf;
            for k=1:numel(indR)
               for l=1:numel(indB)
                  dist=sqrt( (X(indR(k)) - X(indB(l))).^2 + (Y(indR(k)) - Y(indB(l))).^2 + (Z(indR(k)) -
Z(indB(l))).^2 );
                  if dist < minDist, minDist = dist; end
               end
            end
         end
```

```
            borderDistRB(i,j)=minDist;
        end
    end
    toc
    clear indR indB dist minDist X Y Z;

    %% green to blue
    %[X,Y,Z]=meshgrid(1:size(imgG,1),1:size(imgG,2),1:size(imgG,3));
    [X,Y,Z]=meshgrid(0:voxx:(size(imgG,1)-1)*voxx,0:voxy:(size(imgG,2)-1)*voxy,0:voxz:(size(imgG,3)-
1)*voxz);

    borderDistGB=nan(nG,nB);
    tic
    for i=1:nG
        for j=1:nB
            % check if regions are overlapping
            reg1=lmG==i;
            reg2=lmB==j;
            if sum2(and(reg1,reg2))>0
                minDist=0;
                clear reg1 reg2;
            else
                % get voxels belonging to the shells of the current regions
                indG = find(lShellG==i);
                indB = find(lShellB==j);
                % calculate distances of each red shell voxel to each blue shell voxel
                minDist=inf;
                for k=1:numel(indG)
                    for l=1:numel(indB)
                        dist=sqrt( (X(indG(k)) - X(indB(l))).^2 + (Y(indG(k)) - Y(indB(l))).^2 + (Z(indG(k)) -
Z(indB(l))).^2 );
                        if dist < minDist, minDist = dist; end
                    end
                end
            end
            borderDistGB(i,j)=minDist;
        end
    end
    toc
    clear indG indB dist minDist X Y Z;

%% calculate border distances to nearest blob of other color
    % red to green
    closestBorderDistancesRG=min(borderDistRG,[],2);
    % green to red
    closestBorderDistancesGR=min(borderDistRG,[],1);

    % red to blue
    closestBorderDistancesRB=min(borderDistRB,[],2);
    % blue to red
    closestBorderDistancesBR=min(borderDistRB,[],1);

    % green to blue
    closestBorderDistancesGB=min(borderDistGB,[],2);
    % blue to green
    closestBorderDistancesBG=min(borderDistGB,[],1);

        % write table containing overlap
    dlmwrite('closestBorderDistancesGR.txt', closestBorderDistancesGR)
```

```
    dlmwrite('closestBorderDistancesBR.txt', closestBorderDistancesBR)
    dlmwrite('closestBorderDistancesBG.txt', closestBorderDistancesBG)

%% plot histograms of closest neighbor border distances
    subplot(131)
        hist(closestBorderDistancesRG(:),10);
        title('Closest distances red to green')
    subplot(132)
        hist(closestBorderDistancesRB(:),10);
        title('Closest distances red to blue')
    subplot(133)
        hist(closestBorderDistancesGB(:),10);
        title('Closes distances green to blue')


%% determine overlap of neighboring blobs

% red to green
    % find overlapping blobs
    [indR,indG]=find(borderDistRG==0);
    % determine degree of overlap
    overlapRG=nan(size(borderDistRG));
    for i=1:numel(indR)
        region1=lmR==indR(i);
        region2=lmG==indG(i);
        % overlap is defined here as the overlapping area divided by the area of the smaller region
        % it can also be defined as the area_of_intersection / area_of_union * 100, but this does not get
100% even if one
        % region is completely contained in the other one if they are not the same size
        overlapRG(indR(i),indG(i))=sum2( and(region1>0,region2>0) ) / min( sum2(region1>0),
sum2(region2>0) );
    end


% red to blue
    % find overlapping blobs
    [indR,indB]=find(borderDistRB==0);
    % determine degree of overlap
    overlapRB=nan(size(borderDistRB));
    for i=1:numel(indR)
        region1=lmR==indR(i);
        region2=lmB==indB(i);
        % overlap is defined as the overlapping area divided by the area of the smaller region
        overlapRB(indR(i),indB(i))=sum2( and(region1>0,region2>0) ) / min( sum2(region1>0),
sum2(region2>0) );
    end

% green to blue
    % find overlapping blobs
    [indG,indB]=find(borderDistGB==0);
    % determine degree of overlap
    overlapGB=nan(size(borderDistGB));
    for i=1:numel(indG)
        region1=lmG==indG(i);
        region2=lmB==indB(i);
        % overlap is defined as the overlapping area divided by the area of the smaller region
        overlapGB(indG(i),indB(i))=sum2( and(region1>0,region2>0) ) / min( sum2(region1>0),
sum2(region2>0) );
    end
```

```
%% plot overlap histograms to analze distribution
   bins=0.05:0.1:1;

h=bar(bins,[hist(nanmean(overlapRG,1),bins);hist(nanmean(overlapRB,1),bins);hist(nanmean(overlap
GB,1),bins)]');
   xlim([0,1]);
   ylim([0,5]);
   title('overlap of overlapping blobs');
   set(h(1),'FaceColor','r','EdgeColor','g','LineWidth',4);
   set(h(2),'FaceColor','r','EdgeColor','b','LineWidth',4);
   set(h(3),'FaceColor','g','EdgeColor','b','LineWidth',4);
   legend('red to green','red to blue','green to blue','Location','NorthWest');
   shg;


   % calculate mean overlap
   fprintf('\nMean overlap of red and green = %.2f\n',nanmean(overlapRG(:)));
   fprintf('Mean overlap of red and blue = %.2f\n',nanmean(overlapRB(:)));
   fprintf('Mean overlap of green and blue = %.2f\n',nanmean(overlapGB(:)));


   % write table containing overlap
   dlmwrite('overlapGR.txt', nanmean(overlapRG,1))
   dlmwrite('overlapBR.txt', nanmean(overlapRB,1))
   dlmwrite('overlapBG.txt', nanmean(overlapGB,1))

%% plot pixel positions in 3D
   scaleZ=1;
   t=lmR(:,:,:);
   [x,y,z]= ind2sub(size(t),find(t));
   plot3(x,y,z/scaleZ,'ro');
   hold on

   t=lmG(:,:,:);
   [x,y,z]= ind2sub(size(t),find(t));
   plot3(x,y,z/scaleZ,'go'); grid on;

   t=lmB(:,:,:);
   [x,y,z]= ind2sub(size(t),find(t));
   plot3(x,y,z/scaleZ,'bo'); grid on;
   hold off
   axis equal
   shg;

   clear x y z t;

   %% save results to disk

   save(fullfile(dataPath,'results 2.mat'));
```

7.2 Proof of principle image analysis pipeline

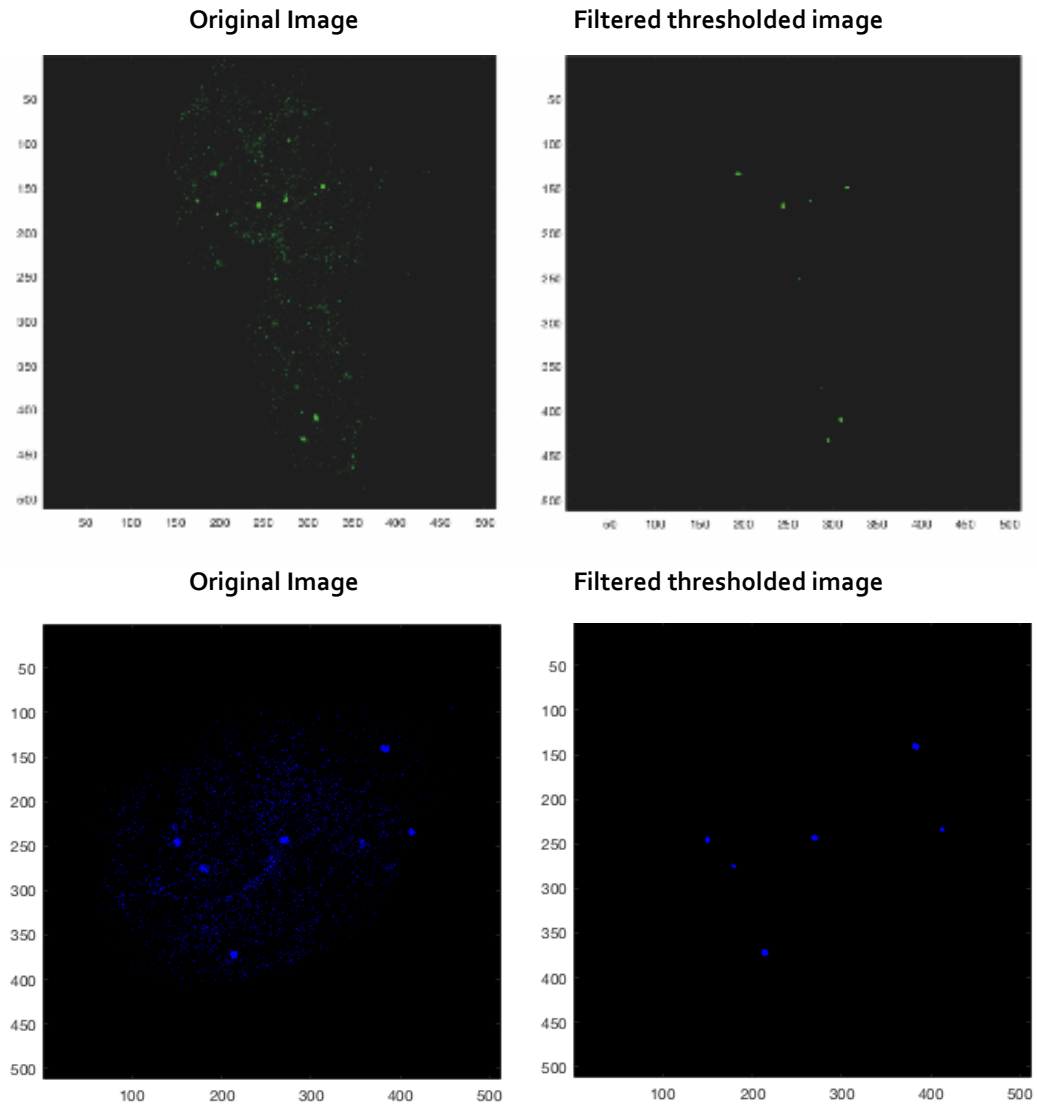


Figure 7. 1: FISH signal detection by automated pipeline
The image on the left shows an optical section of a green channel image. The image on the left shows the selection of signals that are detected by the pipeline, after filtering and thresholding.

7.3 Center of mass cryoFISH distances

The macro was written by Miguel Branco, improved segmentation was added by Tiago Rito.

```
//For single-gene cryo-FISH analysis
//Counts no. of green and red loci in each section, measures their 2D radial position and inter-locus distance
//Measures each section's radius
//In pixels.
//Exports seg image with all the areas detected
//Max entropy for locci, Li thresholding for nuclei

//Input
folder="/ /";
fname="N_";
nStart=1;
nEnd=2;
//f=5

//open("/.tif")


//Start
run("New... ","name=Results type=Table");
print("[Results]","\\Headings:Image no.	Nucleus	Radius Area	nRed	nGreen	R1G1	R1G2	R2G1	R2G2	RadR1 RadR2 RadG1 RadG2");
run("Set Measurements...", "area center");

for (f=nStart; f<=nEnd; f++) {

open(folder+fname+f+".tif");
run("Properties...", "unit=pixel pixel_width=1 pixel_height=1");
rename("RGBim");
run("Duplicate...","title=im");
run("Split Channels");
//run("Tile");
//run("Close All", " ");




//Blue channel
selectWindow("im (blue)");
run("Duplicate...", "title=tmask");
//setAutoThreshold("Li dark");
run("Threshold...");
waitForUser("Threshold", "Please adjust threshold (Li dark recommended) and then click OK");
```

```
run("Make Binary");
run("Fill Holes");
run("Analyze Particles...", "size=40-20000 circularity=0-1 show=Masks exclude clear");
run("Create Selection");
roiManager("reset");
roiManager("add");
close;
selectWindow("im (blue)");
roiManager("select",0);
waitForUser("Correct particles");
roiManager("reset");
roiManager("add");

run("Create Mask");
run("Analyze Particles...", "size=1-Infinity circularity=0-1 show=[Count Masks] clear");
rename("nucmask");
//run("Invert")

print("Num of nuclei:");
print(nResults);

nNuc=nResults;
nucX=newArray(nResults); nucY=newArray(nResults); nucR=newArray(nResults);
nucA=newArray(nResults);//nucC=newArray(nResults);
for (i=0;i<nResults;i++) {
        nucX[i]=getResult("XM",i);
        nucY[i]=getResult("YM",i);
        nucR[i]=sqrt(getResult("Area",i)/PI);
        nucA[i]=getResult("Area",i);
        //nucC[i]=getResult("Circ",i);
}


run("Duplicate...", "title=nucmask8");
run("8-bit");
selectWindow("tmask");
imageCalculator("Subtract", "tmask","nucmask8");
setThreshold(255,255);
run("Create Selection");
roiManager("add");
selectWindow("RGBim");
roiManager("select",1);
run("Fill");
selectWindow("Mask");
close;

print("Nuclei info x,y,r,area:");
for (i=0;i<nNuc;i++) {
        print("Nucleus n.", i);
        print(nucX[i]);
```

```
	print(nucY[i]);
	print(nucR[i]);
	print(nucA[i]);
}




//Red channel

selectWindow("im (red)");
run("Duplicate...", "title=tempR");

run("Threshold...");
//setAutoThreshold("MaxEntropy");
waitForUser("Threshold", "Please adjust threshold (MaxEntropy recommended) and then
click OK");

run("Create Mask"); //run("Invert");
run("Analyze Particles...", "size=2-Infinity circularity=0-1 show=Masks clear");

run("Create Selection");
roiManager("reset");
roiManager("add");
close;
selectWindow("RGBim");
roiManager("select",0);
waitForUser("Correct particles");

print("Num of red:");
print(nResults);

run("Create Mask");
rename("redmask");
run("Analyze Particles...", "size=1-Infinity circularity=0-1 show=Nothing clear");
selectWindow("nucmask");


print("Red loci info x,y,n,r:");

redX=newArray(nResults);
redY=newArray(nResults);
redN=newArray(nResults);
redR=newArray(nResults);

for(i=0;i<nResults;i++) {

	print("Red locus n.", i);
	redX[i]=getResult("XM",i);
```

```
	redY[i]=getResult("YM",i);
	redN[i]=getPixel(redX[i],redY[i]);

	print(redX[i]);
	print(redY[i]);
	print(redN[i]);

	if(redN[i]==0) {
		r=1;
		while (redN[i]==0) {
			makeOval(redX[i]-r,redY[i]-r,2*r,2*r);
			getStatistics(a,m,min,redN[i]);
			r++;
		}
		redR[i]=1;
	} else {
		r=0.5; int=redN[i];
		while (int==redN[i]) {
			makeOval(redX[i]-r,redY[i]-r,2*r,2*r);
			getStatistics(a,int);
			r+=0.5;
		}
		D2C=sqrt(pow(redX[i]-nucX[redN[i]-1],2)+pow(redY[i]-nucY[redN[i]-1],2));
		redR[i]=D2C/(r-0.5+D2C);
	}
	print(redR[i]);
}

//Green loci
selectWindow("im (green)");
run("Duplicate...", "title=tempG");
run("Threshold...");
//setAutoThreshold("MaxEntropy");
waitForUser("Threshold", "Please adjust threshold (MaxEntropy recommended) and then
click OK");
run("Create Mask"); //run("Invert");
run("Analyze Particles...", "size=2-Infinity circularity=0-1 show=Masks clear");
//run("Invert");
run("Create Selection");
roiManager("reset");
roiManager("add");
close;
selectWindow("RGBim");
roiManager("select",0);
waitForUser("Correct particles");

print("Num of green:");
print(nResults);

//if (selectionType==-1) newImage("greenmask", "8-bit White", 512, 512, 1);
```

```
//else
run("Create Mask");
rename("greenmask");//run("Invert");
run("Analyze Particles...", "size=1-Infinity circularity=0-1 show=Nothing clear");
selectWindow("nucmask");

print("Green loci info x,y,n,r:");

greenX=newArray(nResults);
greenY=newArray(nResults);
greenN=newArray(nResults);
greenR=newArray(nResults);

for(i=0;i<nResults;i++) {
	print("Green locus n.", i);
	greenX[i]=getResult("XM",i);
	greenY[i]=getResult("YM",i);
	greenN[i]=getPixel(greenX[i],greenY[i]);

	print(greenX[i]);
	print(greenY[i]);
	print(greenN[i]);


	if(greenN[i]==0) {
		r=1;
		while (greenN[i]==0) {
			makeOval(greenX[i]-r,greenY[i]-r,2*r,2*r);
			getStatistics(a,m,min,greenN[i]);
			r++;
		}
		greenR[i]=1;
	} else {
		r=0.5; int=greenN[i];
		while (int==greenN[i]) {
			makeOval(greenX[i]-r,greenY[i]-r,2*r,2*r);
			getStatistics(a,int);
			r+=0.5;
		}
		D2C=sqrt(pow(greenX[i]-nucX[greenN[i]-1],2)+pow(greenY[i]-
nucY[greenN[i]-1],2));
		greenR[i]=D2C/(r-0.5+D2C);
	}
	print(greenR[i]);
}
```

```
//Segmented image
run("Select None");

selectWindow("redmask"); run("Invert");
selectWindow("greenmask"); run("Invert");
selectWindow("nucmask"); run("Invert");

run("Merge Channels...", "red=redmask green=greenmask blue=nucmask gray=*None*");
run("Invert");
selectWindow("RGB");
saveAs("Tiff", folder+fname+f+"seg.tif");
run("Close All", " ");

//Results
for(i=1;i<=nNuc;i++) {
	nRed=0; radR=newArray(2); whichR=newArray(2);
	for(j=0;j<redN.length;j++) {
		if (redN[j]==i) {
			radR[nRed]=redR[j];
			whichR[nRed]=j;
			nRed++;
		}
	}
	nGreen=0; radG=newArray(2); whichG=newArray(2);
	for(k=0;k<greenN.length;k++) {
		if (greenN[k]==i) {
			radG[nGreen]=greenR[k];
			whichG[nGreen]=k;
			nGreen++;
		}
	}
	dist=newArray(4);
	for(l=0;l<2;l++) {
	if (radR[l]==0) radR[l]="NaN";
		for(m=0;m<2;m++) {
			if (radG[m]==0) radG[m]="NaN";
			if (l>nRed-1||m>nGreen-1) dist[l*2+m]="NaN";
			else dist[l*2+m]=sqrt(pow(redX[whichR[l]]-
greenX[whichG[m]],2)+pow(redY[whichR[l]]-greenY[whichG[m]],2));
		}
	}
	print("[Results]",""+f+"	"+i+"	"+nucR[i-1]+" "+nucA[i-1]+" "+nRed+"
	"+nGreen+"	"+dist[0]+"	"+dist[1]+"	"+dist[2]+"	"+dist[3]+"
	"+radR[0]+"	"+radR[1]+"	"+radG[0]+"	"+radG[1]+"");
}

}
```

7.4 Comparison of Volume acquisition by Matlab and ImageJ pipelines

Comparison of volume analysis with two different approaches: Using an ImageJ based pipeline and a Matlab pipeline. Results show that volumes are comparable, with Matlab results scoring smaller volumes.

Table 7. 1: Comparison Volume acquisition by Matlab and ImageJ

Median Volume 550 ImageJ	Median Volume 550 Matlab	Median Volume 594 ImageJ	Median Volume 594 Matlab
0.818 μm^3	0.868 μm^3	1.003 μm^3	0.973 μm^3

7.5 Improved resolution of Immunofluorescence staining in 3D cells compared to cryosections

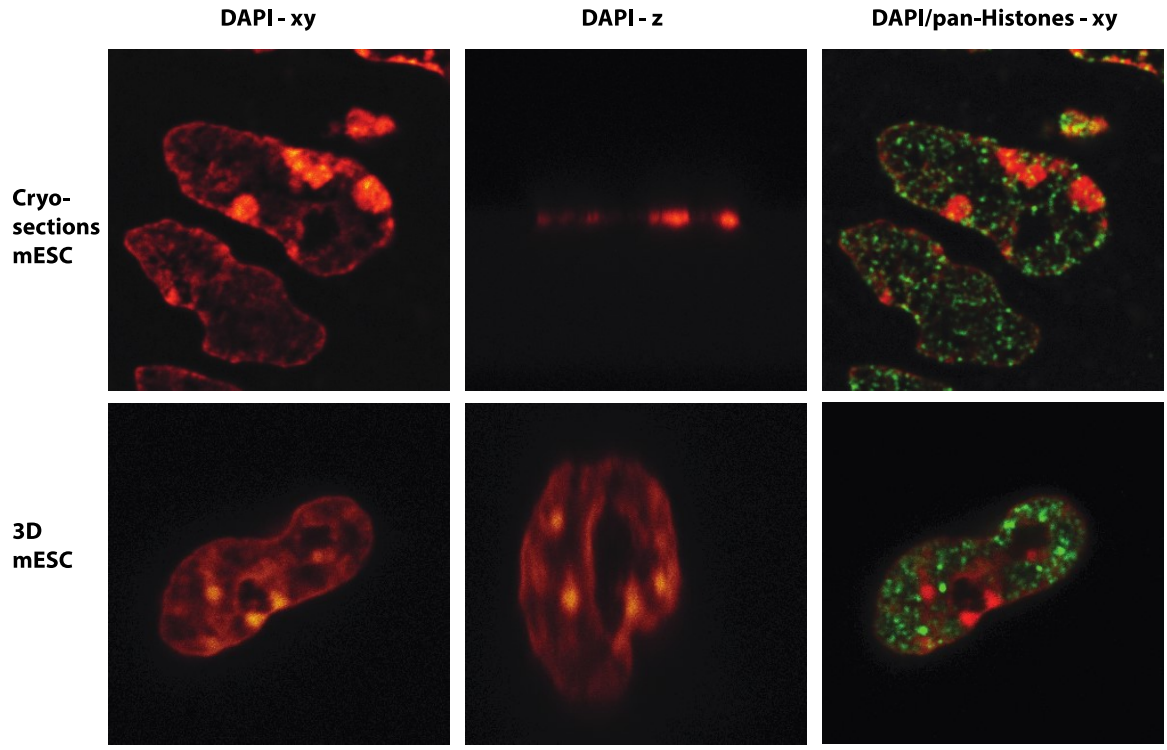


Figure 7.2: Comparison cryosections and 3D ESC
Comparison of DAPI nuclear stain (red) in xy and z, as well as pan-Histones immunofluorescence staining (green) on mESC cryosections and 3D mESC cells. DAPI staining in xy shows better resolution and greater detail in cryosections compared to 3D cells. Cryosections also show chromatin distribution as seen in DAPI staining, with high definition, where many nuclear characteristics such as the nucleolus, chromocenters and chromatin depleted regions are detectable. In 3D cells DAPI staining is more fuzzy and does not reveal the same level of detail. Along the z-axis, cryosections appear as a thin line with a with proportional to section thickness. 3D cells have a decreased resolution in z, however the details of nuclear structure, such as the nucleolus are still discernible. Immunofluorescence staining of pan-Histones has less background signal and higher definition in cryosections, compared to 3D, as no out of focus light is detected.

7.6 STED microscopy on cryosections

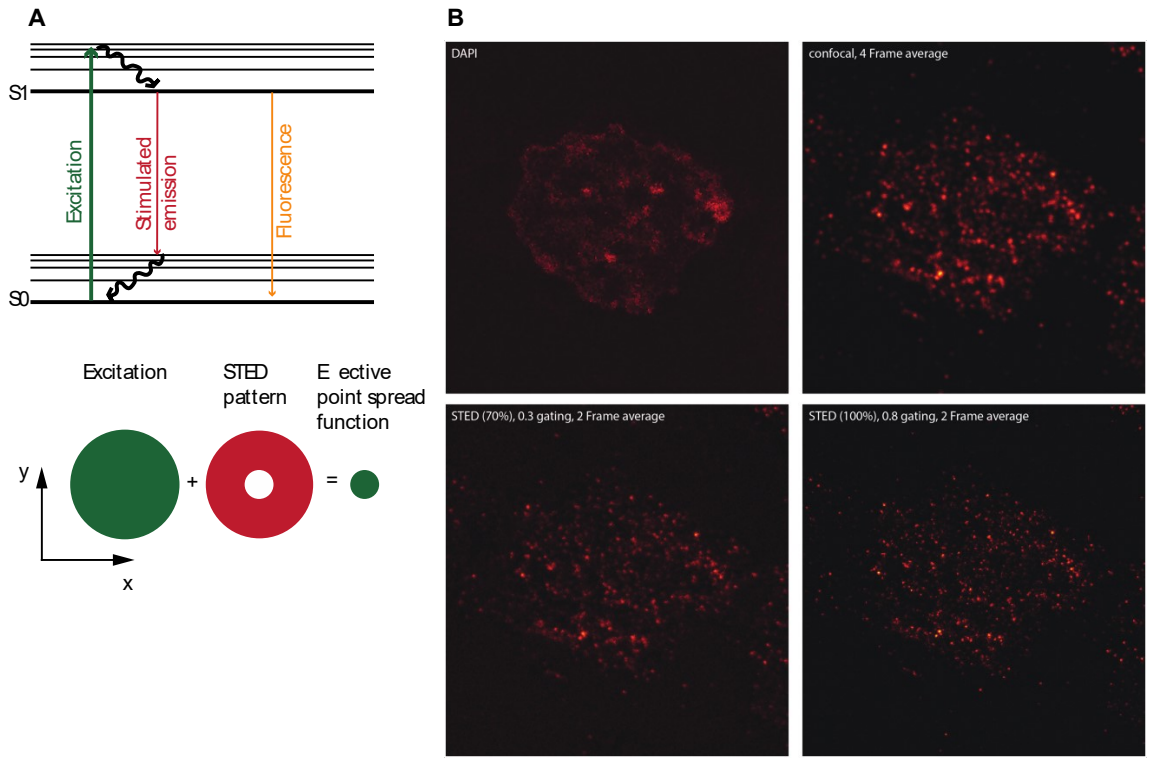
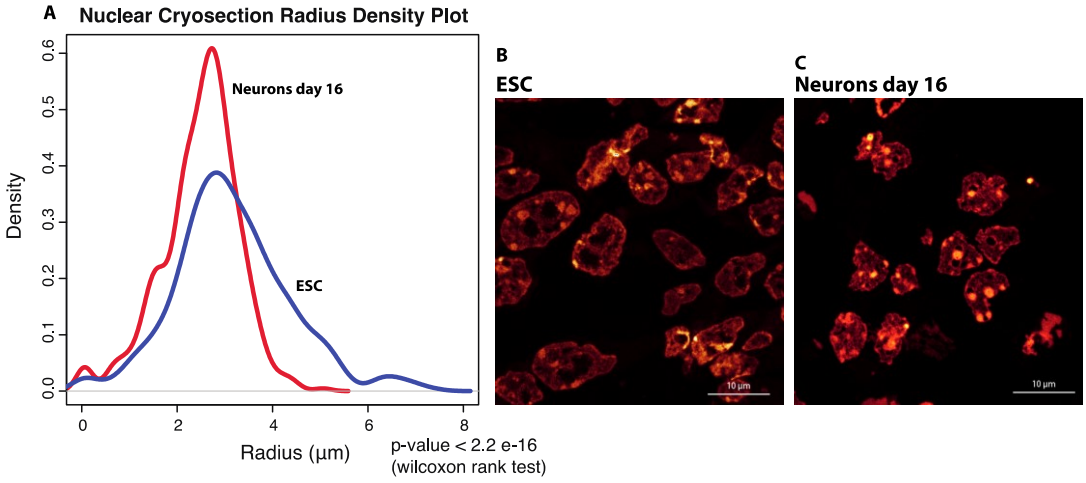


Figure 7.3: Comparison of confocal and STED imaging
A Super-resolution microscopy methods bypass the diffraction limit of light. One such method is stimulated emission depletion microscopy (STED). Super-resolution is achieved by selective deactivation of fluorophores, thereby the area of illumination is minimized at the focal point (Westphal et al., 2008). The Jablonski diagram on the left illustrates the principles of STED. In normal fluorescence, an electron relaxes from its excited electronic state (S1) into the lowest vibrational energy level of its electronic ground state (S0) by releasing a photon that corresponds to the energy gap. STED interferes in this process by forcing the electron to relax into a higher vibrational energy level of the electronic ground state. Since the energy difference in STED forced relaxation is lower, a red-shift in the photon excitation wavelength (corresponding to a longer wavelength) is induced, which later can be distinguished from the core fluorescence signal (Müller et al., 2012). The doughnut shaped STED beam depletes the fluorescence in a specific region, whilst not affecting the center spot, which is free to emit active fluorescence (Török et al., 2004).
B Using stimulated emission depletion (STED) super-resolution imaging, shows that the resolution of immunofluorescence stainings on cryosections can be further improved. The STED resolution depends on optimal imaging settings and adjustments in gating, STED laser beam power translate into increased resolution.

7.7 Nuclear Volumes in ESC and Neurons day 16



	ESC	Neurons	
Average cryosection radius	3.15 µm	2.49 µm	n= 482 (ESC)
Average nuclear cryosection area	50.68 µm²	31.45 µm²	n= 480 (Neurons)
Average nuclear volume	271.37 µm³	132.70 µm³	

Figure 7.4: Nuclear Volumes of ESC and Neurons day 16
A Shows the density plots of nuclear cryosection radii in ESC and Neurons. Neurons have a smaller average radius then ESC. From ESC to Neurons the cryosection area decreases by 38%.
B, C Shows images of DAPI counterstain with cryosectioned nuclei of ESC and Neurons, of the same magnification. Cryosection areas decrease by 38 % from ESC to neurons.

8 Publications

Beagrie, R. A., A. Scialdone, M. Schueler, D. C. Kraemer, M. Chotalia, S. Q. Xie, M. Barbieri, I. de Santiago, L. M. Lavitas, M. R. Branco, J. Fraser, J. Dostie, L. Game, N. Dillon, P. A. Edwards, M. Nicodemi and A. Pombo (2017). "Complex multi-enhancer contacts captured by genome architecture mapping." Nature **543**(7646): 519-524.

Fraser, J., C. Ferrai, A. M. Chiariello, M. Schueler, T. Rito, G. Laudanno, M. Barbieri, B. L. Moore, D. C. Kraemer, S. Aitken, S. Q. Xie, K. J. Morris, M. Itoh, H. Kawaji, I. Jaeger, Y. Hayashizaki, P. Carninci, A. R. Forrest, F. Consortium, C. A. Semple, J. Dostie, A. Pombo and M. Nicodemi (2015). "Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation." Mol Syst Biol **11**(12): 852.

Steimle, S., C. Schnick, E. M. Burger, F. Nuber, D. Kramer, H. Dawitz, S. Brander, B. Matlosz, J. Schafer, K. Maurer, U. Glessner and T. Friedrich (2015). "Cysteine scanning reveals minor local rearrangements of the horizontal helix of respiratory complex I." Mol Microbiol **98**(1): 151-161

9 Figure Permissions

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Apr 05, 2018

This Agreement between Ms. Dorothee Kraemer ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	4320931278583
License date	Apr 02, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature
Licensed Content Title	Formation of new chromatin domains determines pathogenicity of genomic duplications
Licensed Content Author	Martin Franke, Daniel M. Ibrahim, Guillaume Andrey, Wibke Schwarzer, Verena Heinrich et al.
Licensed Content Date	Oct 5, 2016
Licensed Content Volume	538
Licensed Content Issue	7624
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	3
High-res required	yes
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Investigation of Mammalian Chromatin Folding at Different Genomic Length Scales Using High Resolution Imaging
Instructor name	Prof. Ana Pombo
Institution name	Max Delbrück Centre for Molecular Medicine Berlin

Expected presentation date	Apr 2018
Portions	Volume:538, Issue:7624, Pages:265-269, Figure 2, Panel a.
	Same volume, same issue, same pages, Extended Figure 3, Panels a and b.

This Agreement between Ms. Dorothee Kraemer ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	4320920845060
License date	Apr 02, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature
Licensed Content Title	Formation of new chromatin domains determines pathogenicity of genomic duplications
Licensed Content Author	Martin Franke, Daniel M. Ibrahim, Guillaume Andrey, Wibke Schwarzer, Verena Heinrich et al.
Licensed Content Date	Oct 5, 2016
Licensed Content Volume	538
Licensed Content Issue	7624
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	2
High-res required	yes
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Investigation of Mammalian Chromatin Folding at Different Genomic Length Scales Using High Resolution Imaging
Instructor name	Prof. Ana Pombo
Institution name	Max Delbrück Centre for Molecular Medicine Berlin
Expected presentation date	Apr 2018
Portions	Volume:538, Issue:7624, Pages: 265-269, Figure 2, Panel c Same volume, same issue, same pages, Figure 3, Panel a

This Agreement between Ms. Dorothee Kraemer ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	4320930516960
License date	Apr 02, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature
Licensed Content Title	Spatial partitioning of the regulatory landscape of the X-inactivation centre
Licensed Content Author	Elphège P. Nora, Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto et al.
Licensed Content Date	Apr 11, 2012
Licensed Content Volume	485
Licensed Content Issue	7398
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	2
High-res required	yes
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Investigation of Mammalian Chromatin Folding at Different Genomic Length Scales Using High Resolution Imaging
Instructor name	Prof. Ana Pombo
Institution name	Max Delbrück Centre for Molecular Medicine Berlin
Expected presentation date	Apr 2018
Portions	Volume:485, Issue:7398, Pages:381-5, Figure 1, Panels c and e.